PA 1193007

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

July 12, 2004

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE UNDER 35 USC 111.

APPLICATION NUMBER: *60/537,729*
FILING DATE: *January 20, 2004*

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

P. R. GRANT
Certifying Officer

SUBSTITUTE for Provisional Application for Patent Cover Sheet PTO/SB/16 (08-03)
Approved for use through 07/31/2006. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a **PROVISIONAL APPLICATION FOR PATENT** under **37 CFR 1.53 (c)**.

DOCKET NUMBER	ITR0060PV2

INVENTOR(S)

Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Christian	Steinkuhler	Pomezia (Roma), Italy
Armin	Lahm	Pomezia (Roma), Italy
Michele	Pallaoro	Pomezia (Roma), Italy
Caterina	Nardella	Pomezia (Roma), Italy

☐ *Additional inventors are being named on the separately numbered sheets attached hereto*

TITLE OF THE INVENTION (500 characters max)

SYNTHETIC HEPARANASE MOLECULES AND USES THEREOF

CORRESPONDENCE ADDRESS

Direct all Correspondence to:

Merck & Co., Inc.
Patent Department - RY60-30
P.O. Box 2000
Rahway

☒ Customer Number 000210

STATE	New Jersey	ZIP CODE	07065	COUNTRY	U.S.A.

ENCLOSED APPLICATION PARTS *(check all that apply)*

☒ Specification — *Number of Pages* 26
☒ Drawing(s) — *Number of Sheets* 10
☐ Application Data Sheet. See 37 CFR 1.76
☐ CD(s), Number
☒ Other (specify) Sequence Listing (hard copy, 24 pgs)

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT *(check one)*

☐ A check or money order is enclosed to cover the filing fees

☒ The Director is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number: 13-2755

FILING FEE AMOUNT ($) $160.00

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☒ No.

☐ Yes, the name of the U.S. Government agency and the Government contract number are: _____

Respectfully submitted,

SIGNATURE _____

TYPED or PRINTED NAME Alysia A. Finnegan

TELEPHONE 732-594-2588

Date 01/20/2004

REGISTRATION NO. 48,878 *(if appropriate)*

NOTE: Mail to Mail Stop Provisional Application

EXPRESS MAIL CERTIFICATE

DATE OF DEPOSIT January 20, 2004

EXPRESS MAIL NO. EL989603755US

I HEREBY CERTIFY THAT THIS CORRESPONDENCE IS BEING DEPOSITED WITH THE UNITED STATES POSTAL SERVICE AS EXPRESS MAIL "POST OFFICE TO ADDRESSEE" ON THE ABOVE DATE IN AN ENVELOPE ADDRESSED TO COMMISSIONER FOR PATENTS, P.O. BOX 1450, ALEXANDRIA, VA 22313-1450.

MAILED BY _____ **DATE** Jan 20 2004

In Duplicate

PATENT APPLICATION SERIAL NO. _____

U.S. DEPARTMENT OF COMMERCE
PATENT AND TRADEMARK OFFICE
FEE RECORD SHEET

01/23/2004 HGUTEMA1 00000031 132755 60537729
01 FC:1005 160.00 DA

PTO-1556
(5/87)

*U.S. Government Printing Office: 2002 — 489-267/69033

TITLE OF THE INVENTION
SYNTHETIC HEPARANASE MOLECULES AND USES THEREOF

FIELD OF THE INVENTION

The present invention relates to synthetically produced, enzymatically active heparanase molecules that are capable of expression in high yield heterologous expression systems. Also provided herein are methods of expressing mammalian heparanase in heterologous expression systems.

BACKGROUND OF THE INVENTION

Heparan sulfate proteoglycans (HSPGs) are ubiquitous macromolecules found in the extracellular matrix (ECM) and on the cell surface that contribute to the maintenance of cell-cell and cell-ECM interactions. HSPGs are composed of several heparan sulfate (HS) chains covalently linked to a protein core. Heparan sulfate facilitates binding of structural ECM proteins such as fibronectin, laminin, and collagen, to the cell surface and to other ECM proteins, suggesting roles for this glycosaminoglycan in self-assembly and insolubility of ECM components, in cell adhesion, and locomotion. Because of the importance of maintaining proper cell-cell and cell-ECM interactions, HSPGs play crucial structural and regulatory roles in the extracellular milieu, modulating important normal and pathological processes ranging from embryogenesis, morphogenesis and development to inflammation, angiogenesis and cancer metastasis.

In addition to the structural and cell-matrix anchoring roles mentioned above, the structural diversity of HS (Esko et al. *J. Clin. Invest.* 108:169-173 (2001); Turnbull et al. *Trends Cell Biol.* 11: 75-82 (2001)) allows HSPGs to interact with a variety of extracellular signaling proteins such as growth factors, enzymes, and chemokines. Growth factors such as fibroblast growth factors (FGF1 and FGF2), vascular endothelial growth factor (VEGF), hepatocyte growth factor, transforming growth factor β and platelet-derived growth factor, play important roles in tumor growth, invasiveness, and angiogenesis. In addition to acting as a depot for these signaling molecules, activating or stabilizing them, HSPGs may participate in ligand-receptor interactions, such as the binding of FGF2 to the diverse isoforms of the FGF receptor (Chang et al. *FASEB J.* 14: 137-144 (2000)).

Heparan sulfate is degraded by the endo β-*D*-glucuronidase heparanase, which is released by platelets, placental trophoblasts, and leukocytes. Heparanase specifically degrades heparan sulfate by cleaving the glycosidic bond through a hydrolase mechanism. This

degradation results in the release of growth factors such as bFGF, urokinase plasminogen activator (uPA), and tissue plasminogen activator (tPA), which may either initiate neo-angiogenesis or potentiate ECM degradation. Additionally, HS cleavage by heparanase allows cells to migrate through the basal membranes (BM) and traverse the ECM barriers. HS degradation plays an important role in numerous physiological processes by allowing cells to quickly respond to extracellular changes. Therefore, inhibition of heparanase activity could affect pathologies correlated with altered cell migration, such as inflammation, metastasis, and autoimmune disorders.

Due to this pivotal role, heparanase is a potential novel target for the development of antitumor, antimetastasis, or anti-inflammatory drugs. For purposes of drug development, heparanase has a significant advantage over the matrix metalloproteases, which are also ECM-modifying enzymes, because it is likely a single gene product and not part of a complex family of related proteins. Exploiting heparanase as a drug target is presently hampered by both the scarcity of reliable high-throughput assays and by its complex biogenesis, which renders the production of large amounts of active protein a difficult task.

Human heparanase cDNA encodes a protein that is initially synthesized as a pre-pro- protein with a signal peptide sequence that is removed by signal peptidase upon translocation into the endoplasmic reticulum (ER). The resulting 65 kDa pro-form is further processed by removing the 157 N-terminal amino acids to yield the mature 50 kDa heparanase. The 50 kDa protein has a specific activity at least 100 fold higher than the unprocessed 65 kDa precursor (Vlodavsky et al. *Nat. Med.* 5: 793–802 (1999)). Interestingly, the 50 kDa protein is inactive if expressed as such in mammalian cells (Hulett et al. *Nat. Med.* 5: 803–809(1999)). It was proposed that the active form of the enzyme consists of a heterodimer between the 50 kDa fragment and an 8 kDa fragment arising from the excision of an intervening 6 kDa peptide by unidentified proteolytic enzyme(s) (Fairbanks et al. *J. Biol. Chem.* 274: 29587–29590 (1999)). Consistent with this hypothesis, McKenzie et al. (*Biochem J.* 373: 423-435 (2003)) produced active heterodimeric heparanase in insect cells and confirmed that the 8 kDa subunit is necessary for heparanase activity.

Endogenous heparanase can be purified from various sources; however, low heparanase expression levels lead to the necessity for laborious and expensive purification procedures. For example, Toyoshima & Nakajima (*J. Biol. Chem.* 274: 24153-24160 (1999)) described a process for purifying endogenous human heparanase from platelets that requires four different chromatographic steps and lasts five days.

Another drawback to the purification of endogenous heparanase is that overall yields are characteristically low. For instance, Fairbanks *et al.* (*J. Biol. Chem.* 274, 29587-29590, (1999)) report the purification of only 22 µg of heparanase from platelets, with a yield of 6%. Similarly, Fuks and colleagues (U.S. Patent No. 5,362,641) describe a 4000-fold purification of heparanase from 1.4 kg of protein derived from the human hepatoma cell line Sk-Hep-1, producing only 6.5 µg of purified heparanase protein with a yield of 1.9%. A 240,000-fold purification of heparanase from the same cell line was disclosed by Pecker et al. (U.S. Patent No 5,968,822); however, this process required over 500 liters of cell culture.

The identification and cloning of the human heparanase gene (Vlodavksy et al, *Nature Med.* 5: 793-802 (1999); Hulett et al, *Nature Med.* 5: 803-809 (1999); Toyoshima & Nakajima, *J. Biol. Chem.* 274: 24153-24160 (1999)) allowed the recombinant expression of heparanase protein in heterologous expression systems. However, serious deficiencies have been noted with such heterologous expression systems in relation to heparanase production. For example, Ben-Artzi et al. (WO 99/57244) describe the expression of recombinant human heparanase in bacterial, mammalian, yeast, and insect cells. Although heparanase expression was obtained, there was no detectable enzymatic activity associated with the recombinant protein when *E.coli* was host cell, and only the 70 kDa unprocessed precursor was detected when heparanase was expressed in the yeast *Pichia pastoris*.

Ben-Artzi and colleagues (*supra*) also describe the expression of recombinant heparanase in mammalian cells, namely, human kidney fibroblasts (293), baby hamster kidney cells (BHL21) and chinese hamster ovary cells (CHO). However, these expression systems are known to have low yields and high associated costs. Furthermore, despite the fact that processing of the recombinant full-length precursor to yield the active, mature protein is observed in these cells, no homogeneously processed protein is obtained because the processing reaction is inefficient. Additionally, the use of expression vectors driving the secretion of heparanase does not lead to production of recombinant heparanase in the conditioned medium of CHO cells, which have to be further stimulated to secrete heparanase by addition of calcium ionophore or PMA. Only a minor fraction of the secreted protein appeared to be correctly processed in this system.

The production of heparanase in insect cell expression systems such as Sf21 or High five cells is described in the art (WO 99/57244, WO 99/11798, US Patent No. 5,968,822; US Patent No. 6,348,344; and US Patent No. 6,190,875). However, although efficient secretion into the growth medium was observed with such methods, specific activity of the enzyme was very low and no correct processing was observed. For example, Ben-Artzi et al. (WO 99/57244)

describe the introduction of protease cleavage sites downstream of positions 119 or 157 of the heparanase protein in order to generate a correctly processed heparanase in insect cell expression systems. However, these constructs were not shown to be enzymatically active.

McKenzie et al. (*supra*) described the production of active heterodimeric heparanase in insect cells. This system, however, has the disadvantage of requiring the simultaneous production of two different recombinant proteins (the 8 kDa and the 50 kDa subunits). Since admixture of the isolated 8 kDa and 50 kDa domains does not result in heparanase activation, the successful recovery of an active heterodimer by simultaneous expression probably relies on a co-translational formation of the heterodimeric complex. Treatment of this complex with glycanase leads to its dissociation and to the precipitation of the 50 kDa subunit, suggesting a poor stability and solubility.

Despite the methods described above to obtain heparanase in active or inactive form, it would be advantageous to produce biologically active heparanase molecules that are capable of expression in high yield, low cost heterologous expression systems. Said molecules can be used in inhibitor screening assays for the development of therapeutics or pharmaceuticals to inhibit and/or treat metastatic growth and/or inflammation.

SUMMARY OF THE INVENTION

The present invention provides synthetic nucleic acid molecules that encode biologically active, mammalian heparanase, wherein the nucleic acid molecules are capable of expression in high yield heterologous expression systems. The synthetic heparanase molecules provided herein present a significant advance over wild-type heparanase, which is expressed at low levels in mammalian systems and improperly processed in heterologous expression systems. The synthetic molecules of the present invention can be used in inhibitor screening assays for the development of therapeutics or pharmaceuticals to inhibit and/or treat metastatic growth, autoimmune disorders, and/or inflammation.

In one aspect of the invention, the synthetic nucleic acid molecule described above comprises a sequence of nucleotides that encodes a mammalian heparanase protein, the sequence of nucleotides comprising two consensus cleavage sites recognized by an endoproteinase, the cleavage sites located between nucleotides encoding residues 100 and 168 of the heparanase protein. Said nucleic acid molecule encodes a heparanase protein which is capable of biological activity upon incubation with the appropriate enzyme.

This invention further relates to a synthetic mammalian heparanase nucleic acid molecule comprising a portion that encodes a mammalian heparanase protein, the protein coding

portion consisting essentially of a sequence of nucleotides encoding an N-terminal fragment of about 8 kDa, a linker, and a sequence of nucleotides encoding a C-terminal fragment of about 50 kDa, wherein the N-terminal and C-terminal fragments encode protein fragments that are substantially similar to wild-type heparanase fragments and wherein the encoded mammalian heparanase protein is constitutively active.

Also provided herein are synthetically produced, biologically active, mammalian heparanase polypeptides and heparanase polypeptides comprising endoproteinase consensus cleavage sites that are capable of biological activity upon incubation with the appropriate enzyme.

The present invention further provides methods for expressing mammalian heparanase in heterologous expression systems, said methods resulting in high levels of biologically active heparanase expression.

As used throughout the specification and in the appended claims, the singular forms "a," "an," and "the" include the plural reference unless the context clearly dictates otherwise.

As used throughout the specification and appended claims, the following definitions and abbreviations apply:

A "conservative amino acid substitution" refers to the replacement of one amino acid residue by another, chemically similar, amino acid residue. Examples of such conservative substitutions are: substitution of one hydrophobic residue (isoleucine, leucine, valine, or methionine) for another; substitution of one polar residue for another polar residue of the same charge (e.g., arginine for lysine; glutamic acid for aspartic acid).

The term "mammalian" refers to any mammal, including a human being.

The term "treatment" refers to both therapeutic treatment and prophylactic or preventative measures. Those in need of treatment include those already with the disorder as well as those prone to have the disorder or those in which the disorder is to be prevented. A "disorder" is any condition that would benefit from treatment with molecules identified using the nucleic acid molecules and polypeptides described herein. Such disorders include, but are not limited to, cancer, inflammation and autoimmune disorders.

The term "vector" refers to some means by which DNA fragments can be introduced into a host organism or host tissue. There are various types of vectors including plasmid, virus (including adenovirus), bacteriophages and cosmids.

"Biologically active" refers to a protein having structural, regulatory, or biochemical functions attending a naturally occurring molecule or isoform thereof. In the context of heparanase, "biologically active" proteins comprise heparanase enzymatic activity.

"Substantially similar" means that a given sequence shares at least 80%, preferably 90%, more preferably 95%, and even more preferably 99% homology with a reference sequence. In the present invention, the reference sequence can be the full-length human heparanase nucleotide or amino acid sequence, or the nucleotide or amino acid sequence of the 8 kDa (SEQ ID NO:15) or 50 kDa (SEQ ID NO:16) heparanase fragments, as dictated by the context of the text. Thus, a heparanase protein sequence that is "substantially similar" to the 8 kDa human heparanase fragment (SEQ ID NO:15) will share at least 80% homology with the 8 kDa human heparanase fragment, preferably 90% homology, more preferably 95% homology and even more preferably 99% homology. Whether a given heparanase protein or nucleotide sequence is "substantially similar" to a reference sequence can be determined for example, by comparing sequence information using sequence analysis software such as the GAP computer program, version 6.0, available from the University of Wisconsin Genetics Computer Group (UWGCG). The GAP program utilizes the alignment method of Needleman and Wunsch (*J. Mol. Biol.* 48:443, 1970), as revised by Smith and Waterman (*Adv. Appl. Math.* 2:482, 1981).

A "gene" refers to a nucleic acid molecule whose nucleotide sequence codes for a polypeptide molecule. Genes may be uninterrupted sequences of nucleotides or they may include such intervening segments as introns, promoter regions, splicing sites and repetitive sequences. A gene can be either RNA or DNA. A preferred gene is one that encodes the invention peptide.

The term "nucleic acid" or "nucleic acid molecule" is intended for ribonucleic acid (RNA) or deoxyribonucleic acid (DNA), probes, oligonucleotides, fragment or portions thereof, and primers. DNA can be either complementary DNA (cDNA) or genomic DNA, e.g. a gene encoding the invention peptide.

"Wild-type heparanase" or "wild-type protein" or "wt protein" refers to a protein comprising a naturally occurring sequence of amino acids or variant thereof. The amino acid sequence of wild-type human heparanase is available in the art (Vlodavksy et al, *Nature Med.* 5: 793-802 (1999); Hulett et al, *Nature Med.* 5: 803-809 (1999); Toyoshima & Nakajima, *J. Biol. Chem.* 274(34): 24153-24160 (1999); which are herein incorporated by reference in their entirety).

"Wild-type heparanase gene" refers to a gene comprising a sequence of nucleotides that encodes a naturally occurring heparanase protein, including proteins of human

origin or proteins obtained from another organism, including, but not limited to, insects such as *Drosophila*, amphibians such as *Xenopus*, and mammals such as rat, mouse and rhesus monkey. The nucleotide sequence of the human heparanase gene is available in the art (Genbank Accession No. AF155510; Toyoshima and Nakajima, *supra*, which are hereby incorporated by reference in their entirety).

"Substantially free from other proteins" or "substantially purified" means at least 90%, preferably 95%, more preferably 99%, and even more preferably 99.9%, free of other proteins. Thus, a heparanase protein preparation that is substantially free from other proteins will contain, as a percent of its total protein, no more than 10%, preferably no more than 5%, more preferably no more than 1%, and even more preferably no more than 0.1%, of non-heparanase proteins. Whether a given heparanase protein preparation is substantially free from other proteins can be determined by such conventional techniques of assessing protein purity as, *e.g.*, sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) combined with appropriate detection methods, *e.g.*, silver staining or immunoblotting.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1 depicts the biosynthesis of human heparanase in mammalian cells.

FIGURE 2, Panel A, shows a schematic view of the heparanase constructs with engineered TEV cleavage sites. Panel B (left) shows results of Western blot analysis of correctly processed wt heparanase expressed in COS7 cells (lane 1), hepTEV110 (lane 2), hepTEV110 after 16 hours incubation with (lane 3) or without (lane 4) 0.5 μM TEV protease, hepTEV110/158 (lane 5), hepTEV110/158 after 16 hours incubation with (lane 6) or without (lane 7) 0.5 μM TEV protease. Panel B (right) shows heparanase activity of hepTEV110 (column 1), hepTEV110 after 16 hours incubation with (column 2) or without (column 3) 0.5 μM TEV protease, hepTEV110/158 (column 4), hepTEV110/158 after 16 hours incubation with (column 5) or without (column 6) 0.5 μM TEV protease. Heparanase activity of these samples was assessed using the fluorimetric method.

FIGURE 3: Panel A: Multiple sequence alignment of heparanase against related sequences. Predicted secondary structure elements are shown above the alignment (arrows = beta strands, cylinders = helix). The positions of the two cleavage sites are indicated by black triangles. The region of the excised heparanase segment substituted by the Hyaluronidase fragment is surrounded by a grey box. Panel B: Schematic view of the TIM barrel architecture. The location of the excised heparanase segment is indicated with the cleavage points shown as triangles. If present, the segment most likely obscures binding of the substrate (grey arrow) by

several tumors and tumor cell lines. Second, patients with aggressive metastatic disease have measurable heparanase activity in their urine. This observation is not seen with all cancer patients. Additionally, the inhibition of heparanase activity by non-anticoagulant heparin derivatives reduced the incidence of metastases by B16 melanoma, Lewis lung carcinoma and mammary adenocarcinoma cells. Finally, transfection of nonmetastatic murine cells with the human heparanase gene resulted in increased mortality and metastasis in two mouse models.

The human heparanase gene does not share substantial homology with any other known proteins. At the time of its discovery, evidence suggested that heparanase is not a member of a gene family, but rather a single gene or at least the dominant endoglucuronidase involved in HSPG degradation. A second heparanase (hpa2), which shares 34% homology at the amino acid level, was later identified; however, hpa2 seems to serve a different function based on its tissue distribution. The absence of closely related proteins that accomplish analogous tasks, coupled with the above evidence demonstrating a role for heparanase in metastatic growth, make heparanase an excellent target for the development of therapeutics in these areas.

FIGURE 1 depicts the biosynthesis of human heparanase. Briefly, the heparanase cDNA encodes a protein that is initially synthesized as a pre-pro- protein with a signal peptide sequence (residues Met^{1}-Ala^{35}) removed by signal peptidase upon translocation into the ER. The resulting 65 kDa pro-form is further processed by removing the 157 N-terminal amino acids to yield the mature 50 kDa heparanase. The 50 kDa protein has a specific activity at least 100 fold higher than the unprocessed 65 kDa precursor (Vlodavsky et al. *Nat. Med.* 5: 793–802 (1999)). The active form of the enzyme was proposed to be a heterodimer between the 50 kDa fragment and an 8 kDa fragment arising from the excision of an intervening 6 kDa peptide (residues Glu^{109}-Gln^{157}) by unidentified proteolytic enzyme(s) (hereinafter "intervening fragment" or "6 kDa fragment" (Fairbanks et al. *J. Biol. Chem.* 274: 29587–29590 (1999).

Despite recent evidence showing that the 8 kDa subunit is necessary for heparanase activity (McKenzie et al. *Biochem J.* 373: 423-435 (2003)), the role of the 8 kDa subunit in the activation process of heparanase remained unclear prior to the studies disclosed herein: it could function as an essential subunit or, alternatively, act as a chaperone and be dispensable after having accomplished this function. It was also not clear whether other components besides the 8 kDa subunit are necessary to elicit heparanase activation. Multiple sequence alignments and secondary structure prediction lead to a model of the human heparanase according to which the protein adopts a TIM barrel fold, as found in several glycosidases (Hulett et al. *Biochemistry* 39:15659-15667 (2000)). This common fold motif usually consists of 8 alternating α-helices and β-strands. Within the 50 kDa fragment clear homology is observed

beta/alpha units 1 and 2. Design of a shorter loop (dotted line) removes this constraint, leading to an active enzyme while, at the same time, maintaining the structural integrity of the enzyme.

FIGURE 4: Panel A: schematic view of the single chain heparanase constructs described herein. Panel B, left: Western blot analysis of wt heparanase or single chain constructs expressed in COS7 cells. Bla is a control corresponding to the partially purified lysate of COS7 cells transfected only with a vector encoding for the reporter gene β-lactamase (see materials and methods section). Right: Heparanase activity of the same samples using the radiometric assay. Specific activity of all single chain constructs is normalized against that of the wt heparanase.

FIGURE 5: Left, Western blot analysis of the correctly processed wt heparanase produced in COS7 cells or wt heparanase and single chain constructs expressed in Sf9 cells. Right: Heparanase activity of the same samples using the radiometric assay. Specific activity of wt heparanase and single chain constructs expressed in Sf9 cells is normalized against that of the correctly processed wt heparanase produced in COS7 cells.

FIGURE 6: Size exclusion chromatography of FITC-HS degradation products obtained after incubation for 6 hours with hepGS3 (□) and hepHyal (▲) single chain proteins produced in insect cells compared to that of the correctly processed wt heparanase produced in COS7 cells (●) and to unprocessed FITC-HS (○).

FIGURE 7: Ionic strength dependence (panel A), inhibition by heparin (panel B) and pH-dependence (panel C) of wild-type heparanase produced in COS7 cells (●), hepGS3 (□) and hepHyal (▲) single chain constructs produced in insect cells using the fluorimetric activity assay. In the heparin titration experiment, the following IC50 values were obtained: hepwt, 0.9 ng/μl; hepGS3, 1.1 ng/μl; hepHyal, 1.5 ng/μl.

DETAILED DESCRIPTION OF THE INVENTION

Heparanase is a mammalian enzyme that degrades heparan sulfate (HS) by cleaving the glycosidic bond through a hydrolase mechanism. HS degradation plays an important role in numerous physiological processes by allowing cells to quickly respond to extracellular changes by altering cell-cell and cell-ECM interactions. Because of the importance of these interactions, inhibition of heparanase activity could affect several pathologies such as tumor cell metastasis, T-cell mediated delayed type hypersensitivity, and autoimmunity.

Several lines of evidence suggest that heparanase is involved in tumor cell metastasis. First, expression levels of heparanase correlate with the metastatic potential of several tumors and tumor cell lines. Second, patients with aggressive metastatic disease have measurable heparanase activity in their urine. This observation is not seen with all cancer

patients. Additionally, inhibition of heparanase activity by non-anticoagulant heparin derivatives reduced the incidence of metastases by B16 melanoma, Lewis lung carcinoma, and mammary adenocarcinoma cells. Finally, transfection of nonmetastatic murine cells with the human heparanase gene resulted in increased mortality and metastasis in two mouse models.

Human heparanase does not share substantial homology with any other known proteins. At the time of its discovery, evidence suggested that the heparanase gene was not a member of a gene family, but rather a single gene or at least the dominant endoglucuronidase involved in HSPG degradation. A second heparanase (hpa2), which shares 35% identity at the amino acid level, was later identified; however, hpa2 seems to serve a different function based on its tissue distribution. The absence of closely related proteins that accomplish analogous tasks, coupled with the above evidence demonstrating a role for heparanase in metastatic growth, make heparanase an excellent target for the development of therapeutics in these areas.

FIGURE 1 depicts the biosynthesis of human heparanase. Briefly, the heparanase cDNA encodes a protein that is initially synthesized as a pre-pro- protein with a signal peptide sequence (residues Met^{1}-Ala^{35}) removed by signal peptidase upon translocation into the ER. The resulting 65 kDa pro-form is further processed by removing the 157 N-terminal amino acids to yield the mature 50 kDa heparanase (SEQ ID NO:16). The 50 kDa protein has a specific activity at least 100 fold higher than the unprocessed 65 kDa precursor (Vlodavsky et al. *Nat. Med.* 5: 793–802 (1999)). The active form of the enzyme was proposed to be a heterodimer between the 50 kDa fragment and an 8 kDa fragment (SEQ ID NO:15) arising from the excision of an intervening 6 kDa peptide (residues Glu^{109}-Gln^{157}) by unidentified proteolytic enzyme(s) (hereinafter "intervening fragment" or "6 kDa fragment") (Fairbanks et al. *J. Biol. Chem.* 274: 29587–29590 (1999).

Despite recent evidence showing that the 8 kDa subunit (SEQ ID NO:15) is necessary for heparanase activity (McKenzie et al. *Biochem J.* 373: 423-435 (2003)), the role of the 8 kDa subunit in the activation process of heparanase remained unclear prior to the studies disclosed herein: it could function as an essential subunit or, alternatively, act as a chaperone and be dispensable after having accomplished this function. It was also not clear whether other components besides the 8 kDa subunit are necessary to elicit heparanase activation.

Multiple sequence alignments and secondary structure prediction lead to a model of the human heparanase according to which the protein adopts a TIM barrel fold, as found in several glycosidases (Hulett et al. *Biochemistry* 39:15659-15667 (2000)). This common fold motif usually consists of 8 alternating α-helices and β-strands. Within the 50 kDa fragment clear homology is observed only starting with the 3rd α/β unit of the TIM barrel fold, suggesting either

that heparanase adopts a novel fold consisting of only 6 α/β units or that other parts of the protein contribute the missing units. It was postulated that the 8-kDa fragment might contribute the missing structural elements (Hulett et al., *supra*).

Following this hypothesis, a model of the secondary structure of heparanase, based on multiple sequence alignments (FIGURES 3A and 3B), was built to design single chain heparanase molecules having the 8 kDa and the 50 kDa subunits covalently linked together, as described herein. The present invention shows that connecting the 8 kDA and 50kDa fragments with a linker results in constitutively active, single chain heparanase molecules that do not require proteolytic processing. In exemplary embodiments of the invention, the two fragments were connected by grafting of a loop derived from *Hirudinaria manillensis* hyaluronidase or with a linker comprising three glycine-serine repeats.

It is also shown herein that by engineering endoproteinase cleavage sites at about the N and C termini of the 6 kDa intervening fragment, proteolytic processing at both sites of an at least partially purified protein leads to heparanase activation in the absence of other components. In an exemplary embodiment of this aspect of the invention, tobacco etch virus protease cleavage sites are added at the N and C termini of the 6 kDa intervening fragment, resulting in active heparanase after purification or partial purification of the encoded protein and subsequent incubation with the appropriate enzyme. The present invention provides evidence of human heparanase adopting a canonical TIM barrel fold and, advantageously, provides methods for facile production of active enzyme molecules for the identification of specific inhibitors.

The engineered proteins, nucleic acid molecules, and methods of the present invention for expressing biologically active heparanase in heterologous expression systems, particularly insect cells, characteristically produce yields of 0.5 – 5.0 mg/l. Furthermore, these proteins are efficiently secreted into the growth medium, whereas in mammalian cells the authentic human enzyme is mainly retained inside cells or associated with the cell membranes (Vlodavsky et al, *Semin. Cancer Biol.* 12: 121-129 (2002)).

Accordingly, the present invention relates to synthetic nucleic acid molecules that encode an active mammalian heparanase, wherein the nucleic acid molecules are capable of expression in high yield heterologous expression systems. The synthetic heparanase molecules provided herein present a significant advance over wild-type heparanase, which are expressed at low levels in mammalian systems and improperly processed in heterologous expression systems. The synthetic molecules of the present invention can be used in inhibitor screening assays for the development of therapeutics or pharmaceuticals to inhibit and/or treat metastatic growth and/or

inflammation. Said synthetic molecules are also useful in the development of therapeutics or pharmaceuticals for the treatment and/or prevention of autoimmunity.

In one aspect of the present invention, synthetic nucleic acid molecules comprising a sequence of nucleotides that encode a mammalian heparanase protein are provided, the sequence of nucleotides comprising two consensus cleavage sites recognized by an endoproteinase, the cleavage sites located between nucleotides encoding residues 100 and 168 of the heparanase protein. This aspect of the present invention provides synthetic nucleic acid molecules that can be used in methods for carrying out the proteolytic processing of the heparanase protein, similar to the biosynthesis of wild-type heparanase, resulting in a biologically active enzyme.

Also provided herein are substantially pure polypeptides encoded by the nucleic acid molecules described above.

In a preferred embodiment of the invention, the mammalian heparanase protein is human heparanase.

The two consensus cleavage sites can be introduced anywhere between residues 100 and 168 of the heparanase protein, provided that after purification or partial purification of the encoded protein and incubation with the appropriate enzyme, the resulting fragments comprise at least one fragment that is substantially similar to the wild-type 8 kDa fragment (SEQ ID NO:15) and at least one fragment that is substantially similar to the wild-type 50 kDa fragment (SEQ ID NO:16). In a preferred embodiment of the invention, the consensus cleavage sites are located before residues G110 and K158 of the human heparanase protein, resulting in a first fragment of 8 kDa, a second "intervening fragment" of 6 kDa and a third fragment of 50 kDa following purification or partial purification of the encoded protein and subsequent incubation with the appropriate enzyme.

It is understood by one of skill in the art that cleavage sites corresponding to any endoproteinase can be engineered into the heparanase molecule to obtain active, heterodimeric heparanase, including, but not limited to, cleavage sites from tobacco etch virus, 3C protease from picornavirus, thrombin, factor Xa and enterokinase. In a preferred embodiment of the invention, the cleavage sites are from tobacco etch virus.

In another aspect of the present invention, there is provided constitutively active, single-chain mammalian heparanase nucleic acid molecules comprising a portion that encodes a mammalian heparanase protein, the protein coding portion consisting essentially of a sequence of nucleotides encoding an N-terminal fragment of about 8 kDa, a linker, and a sequence of nucleotides encoding a C-terminal fragment of about 50 kDa. This aspect of the present

invention provides synthetic genes encoding heparanase that are constitutively active without proteolytic processing, wherein the synthetic gene is engineered to substantially remove the 6 kDa "intervening fragment" and replace said intervening fragment with a smaller linker.

In preferred embodiments of this aspect of the present invention, the mammalian heparanase protein is a human heparanase.

Also provided herein is a purified synthetic heparanase protein encoded by the constitutively active, single-chain mammalian heparanase gene described above.

Any sequence encoding a peptide comprising from about 1 to about 67 residues can be used as a linker in this aspect of the present invention. Said linker can be synthetic or isolated from a naturally occurring source. In an exemplary embodiment of the present invention, the linker comprises a sequence of nucleotides that encodes a central loop region of the hyaluronidase protein. It is preferred that the hyaluronidase is from *H. manillensis*. In other embodiments, the linker comprises a sequence of nucleotides that encodes a (GlySer)3 linker.

The present invention further relates to recombinant vectors that comprise the synthetic nucleic acid molecules disclosed throughout this specification. These vectors may be comprised of DNA or RNA. For most cloning purposes, DNA vectors are preferred. Typical vectors include plasmids, modified viruses, baculovirus, bacteriophage, cosmids, yeast artificial chromosomes, and other forms of episomal or integrated DNA that can encode a recombinant heparanase protein. It is well within the purview of the skilled artisan to determine an appropriate vector for a particular gene transfer or other use.

An expression vector containing the synthetic nucleic acid molecules disclosed throughout this specification may be used for high-level expression of mammalian heparanase in a recombinant host cell. Expression vectors may include, but are not limited to, cloning vectors, modified cloning vectors, specifically designed plasmids or viruses. Also, a variety of bacterial expression vectors may be used to express recombinant heparanase in bacterial cells if desired. In addition, a variety of fungal cell expression vectors may be used to express recombinant heparanase in fungal cells. Further, a variety of insect cell expression vectors may be used to express recombinant protein in insect cells. In a preferred embodiment of the present invention, the vector is a baculovirus vector.

The present invention also relates to host cells transformed or transfected with vectors comprising the synthetic nucleic acid molecules of the present invention. Recombinant host cells may be prokaryotic or eukaryotic, including but not limited to, bacteria such as *E. coli*, fungal cells such as yeast including, but not limited to, *Pichia pastoris*, *Hansenula polymorpha* and *Saccharomyces cervisiae*, and insect cells, including but not limited to, *Drosophila* and

silkworm derived cell lines. Such recombinant host cells can be cultured under suitable conditions to produce high levels of mammalian heparanase or a biologically equivalent form. As defined herein, the term "host cell" is not intended to include a host cell in the body of a transgenic human being, transgenic human fetus, or transgenic human embryo.

As stated above, the synthetic molecules of the present invention provide a significant advantage over the prior art because they are capable of expression in high-yield heterologous expression systems. The heparanase proteins encoded by the synthetic molecules provided herein are correctly processed, enzymatically active, and expressed to high levels. Therefore, in preferred embodiments of the present invention, the host cell chosen is part of a high yield heterologous expression system, including, but not limited to, insect cells, bacterial cells, and yeast cells. In a particularly preferred embodiment of the present invention, the host cell is an insect cell.

The present invention also relates to recombinant vectors and recombinant host cells, both prokaryotic and eukaryotic, which contain the nucleic acid molecules disclosed throughout this specification. The synthetic nucleic acid molecules, associated vectors, and hosts of the present invention are useful in screening assays to identify inhibitors of heparanase activity, which, are useful for the treatment of cancer, inflammation and/or autoimmunity.

In another aspect of this invention, there is provided a method of expressing mammalian heparanase in non-mammalian cells comprising: (a) transforming or transfecting non-mammalian cells with a vector comprising a sequence of nucleotides that encodes a mammalian heparanase protein, the sequence of nucleotides comprising two consensus cleavage sites recognized by an endoproteinase, the cleavage sites located between nucleotides encoding residues 100 and 168 of the heparanase protein; (b) culturing the host cell under conditions which allow expression of said heparanase protein; (c) disrupting the cells and at least partially purifying the protein; and (d) exposing the at least partially purified protein to the endoproteinase, wherein the heparanase protein is cleaved at the consensus cleavage sites.

This invention also provides substantially purified protein produced by the method described above.

In a preferred embodiment of this aspect of the invention, the mammalian heparanase is human heparanase. In a further preferred embodiment, the consensus cleavage sites are located before residues G110 and K158 of human heparanase.

In another preferred embodiment, the cleavage sites are tobacco etch protein cleavage sites.

Also provided herein is a method of expressing a single chain, constitutively active mammalian heparanase in non-mammalian cells comprising: (a) transforming or transfecting non-mammalian cells with a vector comprising a synthetic mammalian heparanase gene, wherein the synthetic gene comprises a portion that encodes the heparanase protein, the protein coding portion consisting essentially of a sequence of nucleotides encoding an N-terminal fragment of about 8 kDa, a sequence of nucleotides encoding a linker and a sequence of nucleotides encoding a C-terminal fragment of about 50 kDa; and (b) culturing the host cell under conditions which allow expression of said heparanase protein.

Also provided herein is a substantially purified protein produced by the method described above. In a further embodiment of this invention, the protein is capable of binding an antibody that is specific for wild-type heparanase.

In a preferred embodiment of this aspect of the invention, the linker comprises a central loop region of the hyaluronidase protein. In another preferred embodiment, the linker comprises a $(GlySer)_3$ peptide.

All publications mentioned herein are incorporated by reference for the purpose of describing and disclosing methodologies and materials that might be used in connection with the present invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

Having described preferred embodiments of the invention with reference to the accompanying drawings, it is to be understood that the invention is not limited to those precise embodiments, and that various changes and modifications may be effected therein by one skilled in the art without departing from the scope or spirit of the invention as defined in the appended claims.

The following examples illustrate, but do not limit the invention.

EXAMPLE 1

Cloning of heparanase from a human placenta cDNA library.

Human heparanase (Accession No. AF155510) was amplified from a normal human placenta cDNA library (Invitrogen Corp., Carlsbad, CA) by PCR using TaKaRaLa Taq polymerase (TaKaRa Bio Inc., Otsu, Shiga, Japan). Buffer conditions were those suggested by the supplier. PCR amplification of the cDNA templates consisted of one cycle of 94°C for one minute, followed by 35 cycles of 94°C for 30s, 57°C for 30s and 68°C for 110 seconds. The amplified fragment was gel purified, phosphorylated, and cloned either in the *Bam*HI site of

GS4 M1_____W118-(GS)4-E143_____I543

Mutagenic primer: hHEP 329(GS4 Ala) 5' – A C C T T T G A A G A G A G A A G T T A C T G G G G T T C A G G G G C A G G A T C C G G C G C C G A A T G G C C C T A C C A G G A G C A A T T G (SEQ ID NO:7)

Hyaluro M1_____W118-(AFKDKPT) (SEQ ID NO:8)-E143_____I543

Mutagenic primer: hHEP Hyaluro 5' - :A C C T T T G A A G A G A G A A G T T A C T G G G C C T T C A A G G A C A A G A C C C C C G A A T G G C C C T A C C A G G A G C A A T T G – 3' (SEQ ID NO:9)

EXAMPLE 3

Construction of heparanase molecules with engineered protease cleavage sites

To construct an engineered heparanase molecule inserting the consensus cleavage site for the tobacco etch virus (TEV) protease flanked by GS repeats (E109-GSGSENLYFQ-GSG-G110 (SEQ ID NO:10), the scissile bond being located between Q and G) between amino acids E109 and G110, PCR mutagenesis was employed using wt heparanase as a template and the primers hHEP1-24 BamHI opti (SEQ ID NO:1) and hHEP rev 1632 (SEQ ID NO:2) and the mutagenic primer TEV110 bis 5' – G G C A G C G G A T C T G A G A A C C T G T A C T T C C A G G G T T C C G G T T C A A C C T T T G A A G A G A G A A G T T A C – 3' (SEQ ID NO:11).

To construct an engineered heparanase having TEV- cleavage sites both between residues E109/G110 and Q157/K158 the TEV110 construct was used as a template to insert the sequence Q157-GSGSENLYFQ-GSGS-K158 (SEQ ID NO:12) by PCR mutagenesis using the mutagenic primer TEV158 ter 5' – T C T G G A T C C G G T G A A A A T C T C T A T T T T C A G G G C T C A G G A A G T A A A A A G T T C A A G A A C A G C A C C T A C – 3' (SEQ ID NO:13).

All constructs were sequenced on both strands to assure that no mutations were introduced by PCR and cloned into pFASTBAC1 as described above.

EXAMPLE 4

Transient expression of heparanase molecules in COS7 cells.

Cells were grown in Dulbecco's MEM (Gibco BRL, Gaithersburg, MD). All constructs were cloned into the eukaryotic expression plasmid pcDNA3 (Invitrogen). A vector encoding the reporter gene β-lactamase (BLA) was co-transfected in order to check transfection efficiency of each construct. The quantity of each transfected vector was adjusted in order to obtain comparable transfection efficiencies. Transient transfection of COS7 cells was obtained using the fuGENE 6 Transfection Reagent (Roche, Basel, Switzerland) according to manufacturer's instructions. 24 hours after transfection, efficiency was assessed by fluorimetric detection of BLA-positive cells. 96 hours after transfection, cells were harvested and resuspended in lysis buffer (50mM Tris-HCl pH 7.5, 150mM NaCl, 0.5% Triton) containing Complete protease inhibitor cocktail (Roche). The lysis was carried out on ice for 30 minutes. After centrifugation at 14000 rpm for 30 minutes, the heparanase containing supernatants were recovered and partially purified as outlined below.

Heparanase constructs were expressed in COS-7 cells, which are devoid of endogenous heparanse activity, by transient transfection. Heparanase was extracted from cell lysates by heparin affinity chromatography and quantified on Western blots. In parallel, heparanase enzymatic activity was determined with either the radiometric or fluorimetric assay (FIGURE 5). From Western blot analysis, we concluded that wt heparanase as well as the single chain constructs GS3 and hyaluro are efficiently expressed and processed, whereas constructs 106 and GS4 are expressed but not processed. Expression levels of constructs 109 and GS6 were extremely low and barely detectable by Western blot analysis. Only the wt, GS3 and hyaluro constructs showed enzymatic activity. We conclude that single chain constructs 106 and GS4 are inactive whereas constructs 109 and GS6 are probably unstable. Since GS3 and hyaluro are active but are processed despite the changes that were introduced in the cleavage sites we can not draw any conclusion with respect to the intrinsic activity of the precursors. We therefore proceeded with the expression in cells that are devoid of the enzyme(s) responsible for heparanase processing.

EXAMPLE 5

Expression of heparanase molecules in insect cells.

Recombinant baculoviruses containing the heparanase constructs were generated using the Bac to Bac expression system (Invitrogen). Recombinant baculoviruses were used to infect Sf9 insect cells ($50x10^6$ cells per T-175 flask) grown in Grace's insect medium with 10% FBS. Cells were collected 48h after infection, and centrifuged at 500g for 5 minutes. Cell lysates were prepared as above using a lysis buffer with 500mM NaCl instead of 150mM used

A A G T T C A A G A A C A G C A C C T A C – 3' (SEQ ID NO:13), to produce hepTEV110/158 (SEQ ID NOs: 29 and 32).

All constructs were sequenced on both strands to assure that no mutations were introduced by PCR and cloned into pFASTBAC1 as described above.

EXAMPLE 4

Transient expression of heparanase molecules in COS7 cells.

Cells were grown in Dulbecco's MEM (Gibco BRL, Gaithersburg, MD). All constructs were cloned into the eukaryotic expression plasmid pcDNA3 (Invitrogen). A vector encoding the reporter gene β-lactamase (BLA) was co-transfected in order to check transfection efficiency of each construct. The quantity of each transfected vector was adjusted in order to obtain comparable transfection efficiencies. Transient transfection of COS7 cells was obtained using the fuGENE 6 Transfection Reagent (Roche, Basel, Switzerland) according to manufacturer's instructions. 24 hours after transfection, efficiency was assessed by fluorimetric detection of BLA-positive cells. 96 hours after transfection, cells were harvested and resuspended in lysis buffer (50mM Tris-HCl pH 7.5, 150mM NaCl, 0.5% Triton) containing Complete protease inhibitor cocktail (Roche). The lysis was carried out on ice for 30 minutes. After centrifugation at 14000 rpm for 30 minutes, the heparanase containing supernatants were recovered and partially purified as outlined below.

Heparanase constructs were expressed in COS-7 cells, which are devoid of endogenous heparanse activity, by transient transfection. Heparanase was extracted from cell lysates by heparin affinity chromatography and quantified on Western blots. In parallel, heparanase enzymatic activity was determined with either the radiometric or fluorimetric assay (FIGURE 5). From Western blot analysis, we concluded that wt heparanase as well as the single chain constructs GS3 and hyaluro are efficiently expressed and processed, whereas constructs 106 and GS4 are expressed but not processed. Expression levels of constructs 109 and GS6 were extremely low and barely detectable by Western blot analysis. Only the wt, GS3 and hyaluro constructs showed enzymatic activity. We conclude that single chain constructs 106 and GS4 are inactive whereas constructs 109 and GS6 are probably unstable. Since GS3 and hyaluro are active but are processed despite the changes that were introduced in the cleavage sites we can not draw any conclusion with respect to the intrinsic activity of the precursors. We therefore proceeded with the expression in cells that are devoid of the enzyme(s) responsible for heparanase processing.

EXAMPLE 5

Expression of heparanase molecules in insect cells.

Recombinant baculoviruses containing the heparanase constructs were generated using the Bac to Bac expression system (Invitrogen). Recombinant baculoviruses were used to infect Sf9 insect cells ($50x10^6$ cells per T-175 flask) grown in Grace's insect medium with 10% FBS. Cells were collected 48h after infection, and centrifuged at 500g for 5 minutes. Cell lysates were prepared as above using a lysis buffer with 500mM NaCl instead of 150mM used for COS7 (here we used 500mM NaCl because we observed an improvement in protein quantity in the soluble fraction).

The three heparanase constructs that showed enzymatic activity when produced in COS-7 cells were transferred into a baculovirus expression system. The proteins were expressed in Sf9 cells and purified by heparin affinity chromatography. Western blot analysis showed that, in contrast to what was observed in COS-7 cells, no processing of wt or mutant heparanases occurred in this expression system. Analysis of the enzymatic activity of the purified single chain proteins by the fluorimetric activity assay revealed that the unprocessed wt enzyme had a very low activity, whereas the unprocessed GS3 and hyaluro proteins resulted to be highly active, with specific activities comparable to those observed with the correctly processed wild type enzyme produced in COS-7 cells.

GS3 and hyaluro where undistinguishable from the wild type recombinant enzyme extracted from COS-7 cells or from the authentic wt enzyme partially purified from HCT-116 cells on what concerns pH and ionic strength dependence of the enzymatic activity and were inhibited with similar potencies by heparin.

The constructs having TEV cleavage sites at positions 109/110 and 109/110+157/158 were expressed, purified on a heparin affinity column and digested overnight at room temperature with TEV protease (0.5 μM) in 50 mM Mes pH 6.0, 10% glycerol, 0.5 mM EDTA. Complete processing was observed in both cases, however only the double mutant, carrying TEV sequences at both cleavage junctions was activated by this treatment, indicating that processing at the E109/G110 junction only is not sufficient for eliciting activation of heparanase.

EXAMPLE 6

Purification of recombinant heparanase constructs by Heparin Sepharose affinity chromatography.

Cell lysates from COS7 or Sf9 insect cells were passed through 500µl Heparin Sepharose CL-6B (Amersham, Piscataway, NJ) by gravity. The column was washed with 2ml of lysis buffer, then with 2ml of 50mM Tris-HCl pH 7.5, 500mM NaCl, and heparanase was eluted with 2ml of 50mM Tris-HCl pH 7.5, 1 M NaCl and concentrated about 5 fold with a Biomax-30K centrifugal concentrator (Millipore, Bedford, MA). 10% glycerol was added and the protein was stored in aliquots at –80°C. Protein concentration was determined using the BIO-RAD Protein Assay.

EXAMPLE 7

Large scale expression and purification

Sf21 (or Sf9) cells were adapted to growth in serum free medium (Sf-900 II SFM, Invitrogen). Cells were infected with recombinant baculoviruses encoding heparanase constructs at multiplicities of infection varying between 1-10. 3 l of infected cells were grown in spinner flasks at 27°C under a constant flux of sterile air. 48-96 hours after the infection cells were collected and separated from the medium by centrifugation. Synthetic and wt heparanase were found in both the cell pellet and in the supernatant. To extract synthetic heparanase from the cell pellet, cells were disrupted as outlined above. Cell lysates or the crude medium supernatant were filtered on a 0.22µ filter and loaded on a 20 ml-HyperD Heparin column (Biosepra Inc., Marlboro, MA) equilibrated with 50 mM Tris-HCl pH 7.5, 150 mM NaCl. Synthetic or wt heparanase were eluted by applying a linear 0.15- 1M NaCl gradient in 50 mM Tris HCl pH 7.5. Recombinant proteins eluted at NaCl concentrations >500 mM. The pooled, heparanase-containing Heparin-column fractions were dialyzed overnight against 50 mM HEPES pH 7.5 and loaded on a Source S column (Amersham) equilibrated in the same buffer. Heparanase constructs eluted with 400-600 mM NaCl. Proteins were purified to homogeneity by a further chromatographic step on a 15/30 Superdex 75 size exclusion column. The purified proteins were aliquoted, shock-frozen in liquid nitrogen and stored at -80 °C.

EXAMPLE 8

Western Blotting.

Rabbit polyclonal antibodies were generated against a peptide contained within the 50 kDa subunit (EPNSFLKKADIFINGSQ (SEQ ID NO:14), corresponding to amino acids 225 to 241 and containing the additional sequence GGC at its C-terminus). Antisera were immunopurified using the immunogen peptide immobilized on a thiopropyl Sepharose resin (Amersham). 10µl of proteins eluted from the heparin column were subjected to 10% SDS-

polyacrylamide gel electrophoresis and transferred onto Protran BA 83 Cellulosenitrate membrane (Schleicher & Schuell Bioscience, Keene, NH). After saturation of non specific binding with 5% milk, the membrane was incubated with the polyclonal antibody described above diluted 1:500 in 5% milk, TBS and 0.05% Tween20 over night at 4°C. After washing, the membrane was incubated with anti-rabbit horseradish peroxidase-conjugated antibody diluted 1:5000 for 30' at room temperature. The immunoreactive bands were detected by SuperSignal West Pico Chemiluminescent Substrate (Pierce Biotechnology, Rockford, IL). Finally the membrane was exposed to BIOMAX MR film (Kodak) for 10s.

EXAMPLE 9

Fluorometric labeling of heparan sulfate

Heparan sulfate sodium salt from bovine kidney (Sigma-Aldrich Corp., St. Louis, MO) was labeled with fluorescein isothiocyanate (FITC) as previously described (Toyoshima and Nakajima, *J. Biol. Chem.* 274: 24153-24160 (1999)). 5 mg of heparan sulfate and 5 mg of FITC were dissolved in 1 ml of 0.1 M Na_2CO_3 pH 9.5 and incubated over night at 4°C in the dark. The solution was then loaded on MicroSpin G-25 columns in order to separate FITC labeled Heparan Sulfate (FITC-HS) from unreacted FITC. The FITC-HS was subjected to a first gel-filtration chromatographic step through Sephacryl S-300 in 150mM NaCl, 25mM Tris-HCl pH=7.5 buffer to separate the high molecular weight heparan sulfate species. The colored fractions were pooled, concentrated with Biomax-10K centrifugal concentrator (Millipore) and rechromatographed on Sephacryl S-300 (as above) in order to obtain heparan sulfate species with homogeneous molecular weight. The eluted fractions were analyzed by HPLC Superdex 75™ (Pharmacia Biotech) chromatography system. The fluorescence in each fraction was measured by an L-7485 fluorescence detector (Merck Hitachi). We obtained four main fractions with different molecular weight heparan sulfate products. The quantity of FITC-HS in each fraction was measured with the Blyscan Glycosaminoglycan Assay (Biocolor Ltd., Belfast, Northern Ireland).

EXAMPLE 10

Fluorimetric assay.

This assay is based on the degradation of FITC-HS monitored by HPLC size exclusion chromatography. 8µl of purified heparanase was incubated with 5µl of FITC-HS in a 50µl of 50 mM MES pH 6, 10% glycerol (heparanase activity buffer, HAB). The reaction mixture was incubated at room temperature for a defined period and the reaction was stopped by

the addition of 50µg of heparin. The mixture was then filtered using Ultrafree-MC centrifugal filter Devices (Millipore). 20µl were injected on a Superdex 75™ (Pharmacia Biotech) column equilibrated in buffer 50mM Hepes pH 7.5 150mM Na_2SO_4 and connected to a Merck-Hitachi HPLC system. Fluorescent heparan sulfate degradation products were detected by an L-7485 fluorescence detector. Heparanase activity was assessed by monitoring the increase in lower molecular weight heparan sulfate species compared with the intact FITC-HS and quantified by peak area integration.

EXAMPLE 11

Radiometric labeling and biotinylation at the reducing end of heparan sulfate

10mg of heparan sulfate sodium salt from bovine kidney (Sigma) were partially N-de-acetylated and re-acetylated with [^{3}H] acetic anhydride as previously described (Freeman and Parish, *Biochem. J.* 325: 229-237 (1997)). Tritiated heparan sulfate was then subjected to reductive amination at the reducing end as described. Tritiated, reductively aminated heparan sulfate was further conjugated to biotin using EZ-Link Sulfo-NHS-LC-Biotin (Pierce). This biotin analog has an N-hydroxysuccinimido ester moiety that can react with the amino group generated at the reducing end of the heparan sulfate molecules. We calculated a recovery in about 5mg of tritiated heparan sulfate, reductively aminated and resuspended in 1 ml of H_2O (an estimated final concentration of 100 µM taking into account an average in heparan sulfate molecular weight of 500KDa). To 100µl of this solution 1mg of EZ-Link Sulfo-NHS-LC-Biotin (about 100-fold molar excess) and 20µl of phosphate buffer pH 7.5 were added. The reaction mixture was incubated overnight at room temperature. The reaction mixture was then loaded on PD-10 desalting column in order to separate biotinylated, tritiated heparan sulfate from unreacted biotin. We finally obtained four fractions (1 ml each), which were tested for their ability to be immobilized on Reacti-Bind Streptavidin High Binding Capacity Coated Plates (Pierce).

EXAMPLE 12

Radiometric assay.

This assay is based on the degradation of tritiated heparan sulfate immobilized on microplate. Each well of the Reacti-Bind Streptavidin High Binding Capacity Coated Plates was pre-treated according to manufacturer's instructions. Initially, different amounts of each fraction of tritiated, biotinylated heparan sulfate obtained after PD-10 desalting column were added to each well (in duplicate) in PBS to a final volume of 100µl. After assesseing that the maximum

binding is obtained with a volume of fraction 2 corresponding to $100x10^3$ d.p.m. this amount was always used. The binding was carried out over night at room temperature. The wells were then washed three times with PBS and twice with HAB. 10µl of purified heparanase were added to each well in HAB to a final volume of 100µl. The wells were incubated at room temperature for 2-24 hours. Finally, the liberated radioactivity due to tritiated heparan sulfate products generated by heparanase in each well was measured and normalized against a buffer blank.

EXAMPLE 13

Determination of specific activity of heparanase constructs.

Specific activities of the heparanase constructs either transiently expressed in COS7 cells or expressed in the baculovirus system were determined as follows:

$$\text{Specific activity} = \frac{\text{normalized activity (d.p.m./µl)}}{\text{normalized densitometric volume (volume/µl)}}$$

In detail, activity of partially purified heparanse constructs was determined in the radiometric assay by titrating each preparation in such a way that a linear dose-activity relationship was observed. These titrations were repeated three times with each preparation and a mean, normalized activity (d.p.m./µl) was calculated. Protein expression was determined by the Western blotting experiments: the chemiluminescent readout was quantified by densitometry. Again, experiments were repeated three times and mean values were determined. The specific activity was obtained by dividing the normalized activity (d.p.m./µl) by the normalized densitometric volume (volume/µl).

WHAT IS CLAIMED IS:

1. A synthetic nucleic acid molecule comprising a sequence of nucleotides that encodes a mammalian heparanase protein, the sequence of nucleotides comprising two consensus cleavage sites recognized by an endoproteinase, the cleavage sites located between nucleotides encoding residues 100 and 168 of the heparanase protein.

2. A vector comprising the nucleic acid molecule of claim 1.

3. The vector of claim 2, wherein the vector is a baculovirus vector.

4. A host cell comprising the vector of claim 3.

5. The host cell of claim 4, wherein the host cell is an insect cell.

6. The host cell of claim 4, wherein the host cell is a yeast cell.

7. The host cell of claim 6, wherein the yeast is selected from the group consisting of: *Pichia pastoris*, *Hansenula polymorpha* and *Saccharomyces cervisiae*.

8. The synthetic nucleic acid molecule of claim 1, wherein the heparanase protein is human heparanase.

9. The synthetic nucleic acid molecule of claim 8, wherein the consensus cleavage sites are located before residues G110 and K158 of the human heparanase protein.

10. The synthetic nucleic acid molecule of claim 8, wherein the consensus cleavage sites are selected from the group consisting of: tobacco etch virus (TEV) protease cleavage sites, 3C protease cleavage sites from picornavirus, thrombin protease cleavage sites, enterokinase cleavage sites and factor Xa cleavage sites.

11. A synthetic mammalian heparanase nucleic acid molecule comprising a portion that encodes a mammalian heparanase protein, the protein coding portion consisting essentially of a sequence of nucleotides encoding an N-terminal fragment of about 8 kDa, a

linker, and a sequence of nucleotides encoding a C-terminal fragment of about 50 kDa, wherein the N-terminal and C-terminal fragments encode protein fragments that are substantially similar to wild-type heparanase fragments, and wherein the encoded heparanase protein is constitutively active.

12. The gene of claim 11, wherein the protein coding portion encodes human heparanase.

13. The gene of claim 11, wherein the linker comprises a sequence of nucleotides that encodes a central loop region of the hyaluronidase protein.

14. The gene of claim 13, wherein the hyaluronidase is from *H. manillensis*.

15. The gene of claim 12, wherein the linker comprises a sequence of nucleotides that encodes a $(GlySer)_3$ linker.

16. A vector comprising the gene of claim 12.

17. A host cell comprising the vector of claim 16.

18. The host cell of claim 17 which is an insect cell or a yeast cell.

19. A purified synthetic heparanase protein encoded by the gene of claim 12.

20. A method of expressing mammalian heparanase in non-mammalian cells comprising:

(a) transforming or transfecting non-mammalian cells with a vector comprising a sequence of nucleotides that encodes a mammalian heparanase protein, the sequence of nucleotides comprising two consensus cleavage sites recognized by an endoproteinase, the cleavage sites located between residues 100 and 168 of the heparanase protein;

(b) culturing the host cell under conditions which allow expression of said heparanase protein;

(c) disrupting the cells and at least partially purifying the heparanase protein; and

(c) exposing the at least partially purified heparanase protein to the endoproteinase, wherein the heparanase protein is cleaved at the consensus cleavage sites.

21. A method as in claim 20, wherein the heparanase is human.

22. A method of expressing a single chain, constitutively active mammalian heparanase in non-mammalian cells comprising:

(a) transforming or transfecting non-mammalian cells with a vector comprising a synthetic mammalian heparanase gene, wherein the synthetic gene comprises a portion that encodes the heparanase protein, the protein coding portion consisting essentially of a sequence of nucleotides encoding an N-terminal fragment of about 8 kDa, a sequence of nucleotides encoding a linker and a sequence of nucleotides encoding a C-terminal fragment of about 50 kDa, wherein the N-terminal and C-terminal fragments encode protein fragments that are substantially similar to wild-type fragments; and

(b) culturing the host cell under conditions which allow expression of said heparanase protein

23. The method of claim 22 wherein the linker comprises a central loop region of the hyaluronidase protein.

24. The method of claim 22 wherein the linker comprises a central (GlySer)3.

25. A substantially pure protein produced by the method of claim 22.

TITLE OF THE INVENTION
SYNTHETIC HEPARANASE MOLECULES AND USES THEREOF

ABSTRACT OF THE DISCLOSURE

The present invention relates to synthetically produced, enzymatically active heparanase nucleic acid molecules that are capable of expression in high yield heterologous expression systems, and to polypeptides encoded by said molecules. Also provided herein are methods of expressing mammalian heparanase in heterologous expression systems, wherein high yields of biologically active heparanase are produced compared to prior art methods.

FIGURE 1

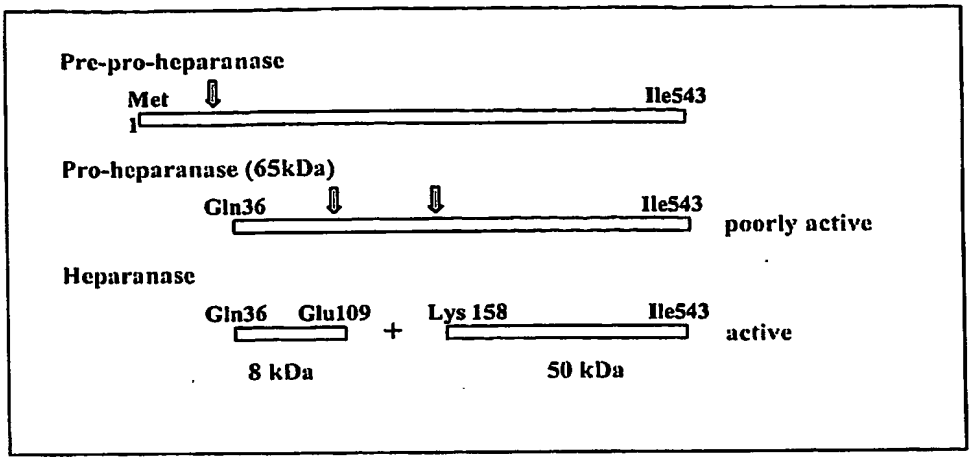
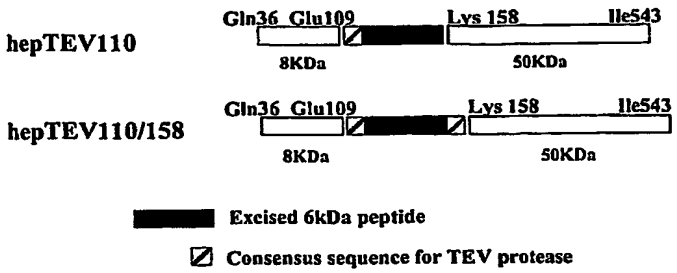

Pre-pro-heparanase
Met
1
Ile543
Pro-heparanase (65kDa)
Gln36
Ile543
poorly active
Heparanase
Gln36
Glu109
+
Lys 158
Ile543
active
8 kDa
50 kDa

FIGURE 2

A.


hepTEV110
Gln36 Glu109
Lys 158
Ile543
8KDa
50KDa
hepTEV110/158
Gln36 Glu109
Lys 158
Ile543
8KDa
50KDa
Excised 6kDa peptide
Consensus sequence for TEV protease

B.

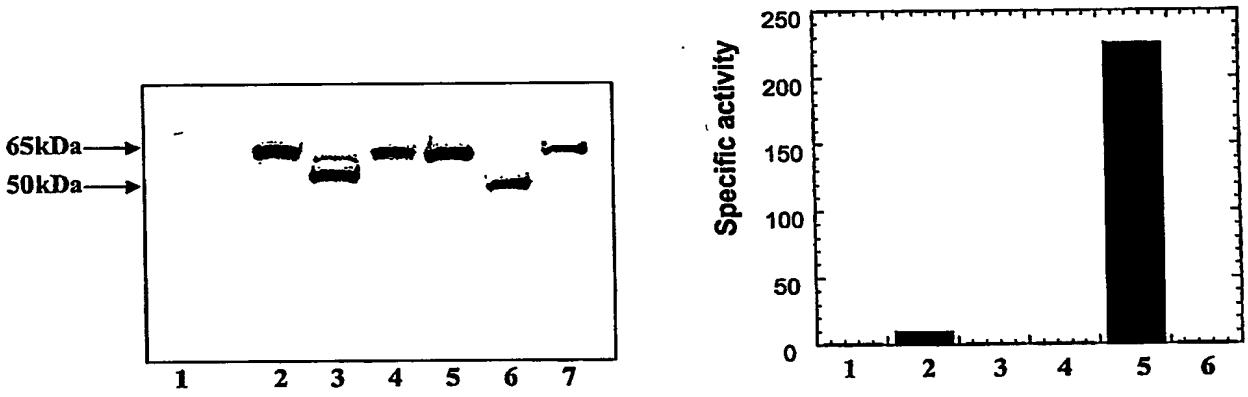

65kDa
50kDa
1
2
3
4
5
6
7
Specific activity
250
200
150
100
50
0
1
2
3
4
5
6

FIGURE 3A

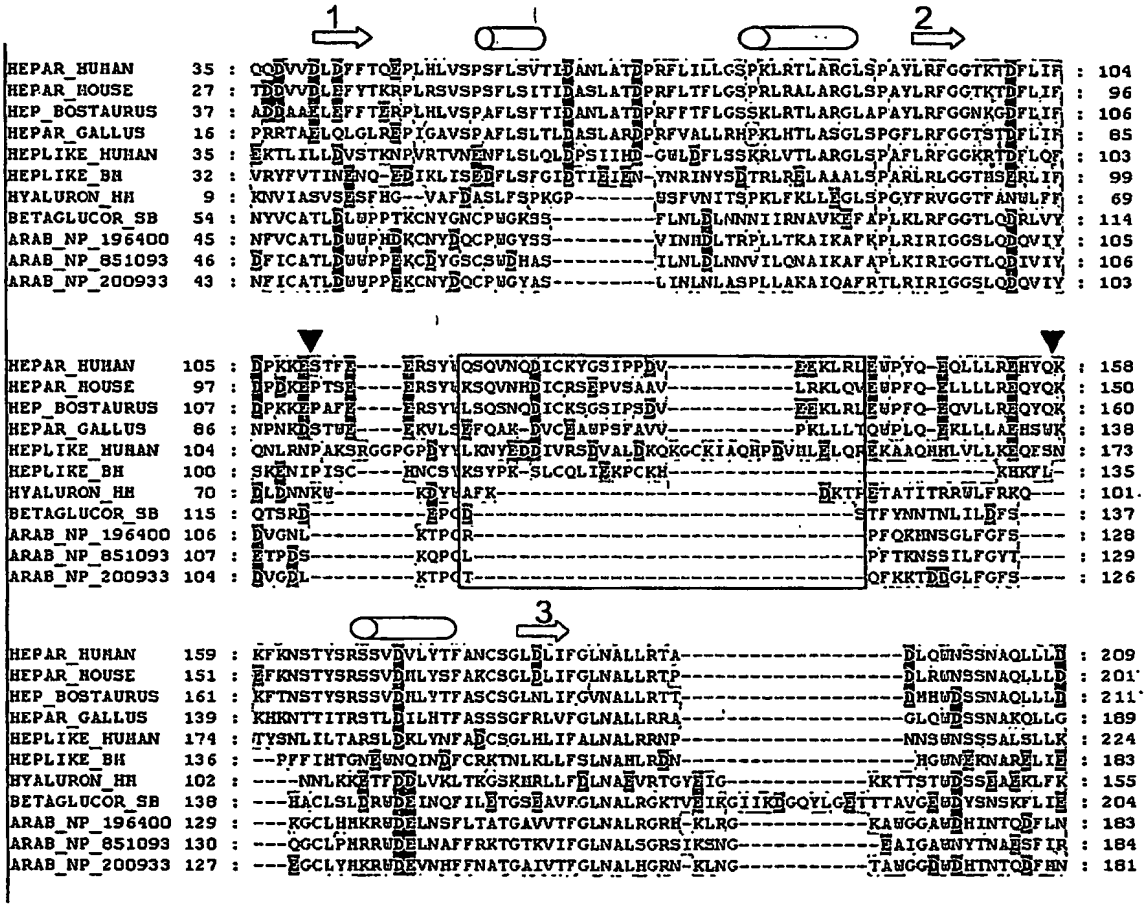

1
2
3
HEPAR_HUMAN 35 : QQDVVDLDFFTQEPLHLVSPSFLSVTIDANLATDPRFLILLGSPKLRTLARGLSPAYLRFGGTKTDFLIF : 104
HEPAR_MOUSE 27 : TDDVVDLEFYTKRPLRSVSPSFLSITIDASLATDPRFLTFLGSPRLRALARGLSPAYLRFGGTKTDFLIF : 96
HEP_BOSTAURUS 37 : ADDAAELEFFTERPLHLVSPAFLSFTIDANLATDPRFFTFLGSSKLRTLARGLAPAYLRFGGNKGDFLIF : 106
HEPAR_GALLUS 16 : PRRTAELQLGLREPIGAVSPAFLSLTLDASLARDPRFVALLRHPKLHTLASGLSPGFLRFGGTSTDFLIF : 85
HEPLIKE_HUMAN 35 : EKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHD-GWLDFLSSKRLVTLARGLSPAFLRFGGKRTDFLQF : 103
HEPLIKE_BH 32 : VRYFVTINENQ-EDIKLISEDFLSFGIDTIEIEN-YNRINYSDTRLRELAAALSPARLRLGGTHSERLIF : 99
HYALURON_HH 9 : KNVIASVSESFHG--VAFDASLFSPKGP--------GSFVNITSPKLFKLLEGLSPGYFRVGGTFANWLFF : 69
BETAGLUCOR_SB 54 : NYVCATLDLGPPTKCNYGNCPWGKSS---------FLNLDLNNNIIRNAVKEFAPLKLRFGGTLQDRLVY : 114
ARAB_NP_196400 45 : NFVCATLDWGPHDKCNYDQCPWGYSS---------VINHDLTRPLLTKAIKAFKPLRIRIGGSLQDQVIY : 105
ARAB_NP_851093 46 : DFICATLDWGPPEKCDYGSCSWDHAS---------ILNLDLNNVILQNAIKAFAPLKIRIGGTLQDIVIY : 106
ARAB_NP_200933 43 : NFICATLDWGPPEKCNYDQCPWGYAS---------LINLNLASPLLAKAIQAFRTLRIRIGGSLQDQVIY : 103
HEPAR_HUMAN 105 : DPKKESTFE----ERSYLQSQVNQDICKYGSIPPDV----------EEKLRLEWPYQ-EQLLLREHYQK : 158
HEPAR_MOUSE 97 : DPDKEPTSE----ERSYLKSQVNHDICRSEPVSAAV----------LRKLQVEWPFQ-ELLLLREQYQK : 150
HEP_BOSTAURUS 107 : DPKKEPAFE----ERSYLLSQSNQDICKSGSIPSDV----------EEKLRLEWPFQ-EQVLLREQYQK : 160
HEPAR_GALLUS 86 : NPNKDSTWE----EKVLSEFQAK-DVCEAWPSFAVV----------PKLLLTQWPLQ-EKLLLAEHSWK : 138
HEPLIKE_HUMAN 104 : QNLRNPAKSRGGPGPDYLKNYEDDIVRSDVALDKQKGCKIAQHPDVHLELQPEKAAQHHLVLLKEQFSN : 173
HEPLIKE_BH 100 : SKENIPISC----HNCSVKSYPK-SLCQLIEKPCKH----------KHKFL- : 135
HYALURON_HH 70 : ELDNNKW--------KDYLAFK----------DKTPETATITRRWLFRKQ--- : 101
BETAGLUCOR_SB 115 : QTSRD----------EPCD----------STFYNNTNLILDFS---- : 137
ARAB_NP_196400 106 : DVGNL---------KTPCR----------PFQKHNSGLFGFS---- : 128
ARAB_NP_851093 107 : ETPDS---------KQPCL----------PFTKNSSILFGYT---- : 129
ARAB_NP_200933 104 : DVGDL---------KTPCT----------QFKKTDDGLFGFS---- : 126
HEPAR_HUMAN 159 : KFKNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTA----------DLQWNSSNAQLLLD : 209
HEPAR_MOUSE 151 : EFKNSTYSRSSVDHLYSFAKCSGLDLIFGLNALLRTP----------DLRWNSSNAQLLLD : 201
HEP_BOSTAURUS 161 : KFTNSTYSRSSVDHLYTFASCSGLNLIFGVNALLRTT----------DMHWDSSNAQLLLD : 211
HEPAR_GALLUS 139 : KHKNTTITRSTLDILHTFASSSGFRLVFGLNALLRRA----------GLQWDSSNAKQLLG : 189
HEPLIKE_HUMAN 174 : TYSNLILTARSLDKLYNFADCSGLHLIFALNALRRNP----------NNSWNSSSALSLLK : 224
HEPLIKE_BH 136 : --PFFIHTGNEWNQINDFCRKTNLKLLFSLNAHLRDN----------HGWNEKNARELIE : 183
HYALURON_HH 102 : -----NNLKKETFDDLVKLTKGSKHRLLFDLNAEVRTGYEIG----------KKTTSTWDSSHAEKLFK : 155
BETAGLUCOR_SB 138 : ---HACLSLDRWDEINQFILETGSEAVFGLNALRGKTVEIKGIIKDGQYLGETTTAVGEWDYSNSKFLIE : 204
ARAB_NP_196400 129 : ---KGCLHHKRWDELNSFLTATGAVVTFGLNALRGRH-KLRG----------KAWGGAWDHINTQDFLN : 183
ARAB_NP_851093 130 : ---QGCLPHRRWDELNAFFRKTGTKVIFGLNALSGRSIKSNG----------EAIGAWNYTNAESFIR : 184
ARAB_NP_200933 127 : ---EGCLYHKRWDEVNHFFNATGAIVTFGLNALHGRN-KLNG----------TAWGGEWDHINTQDFHN : 181

FIGURE 3B

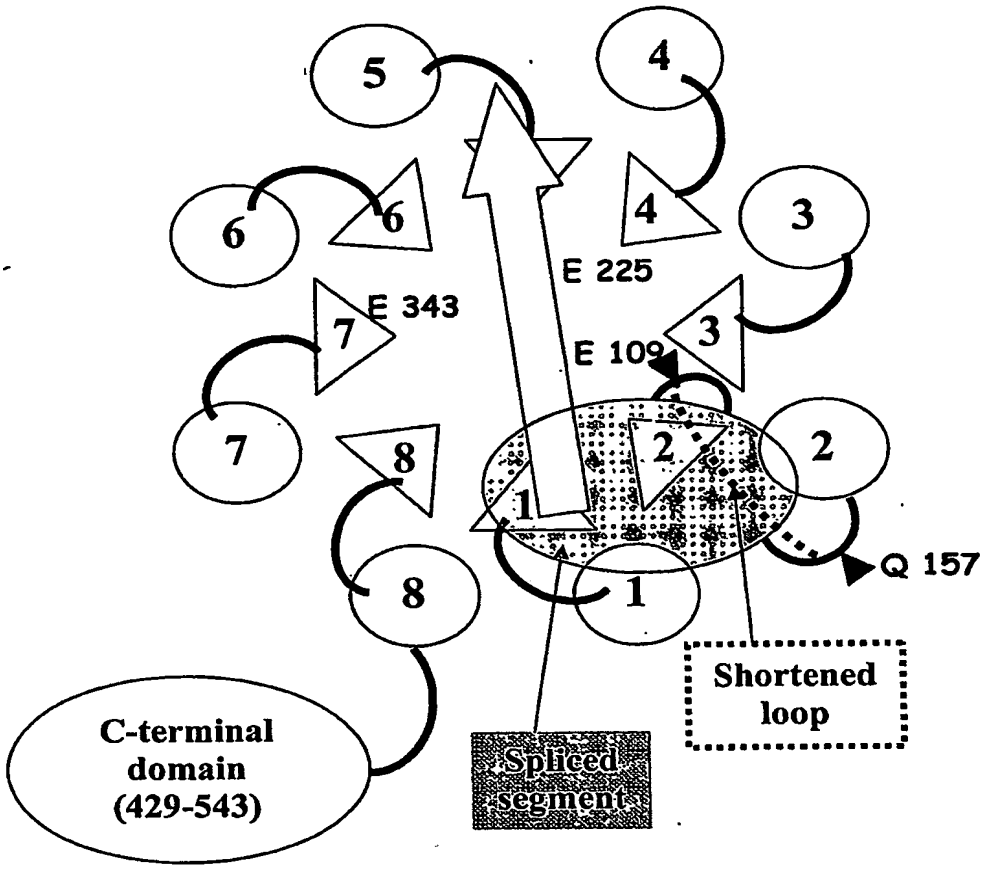
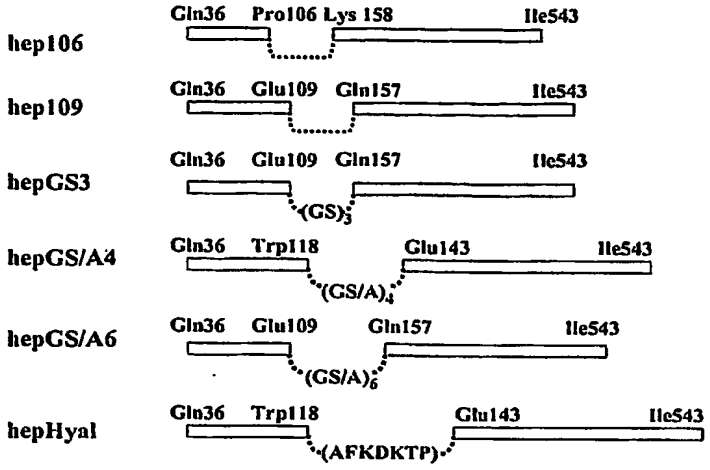
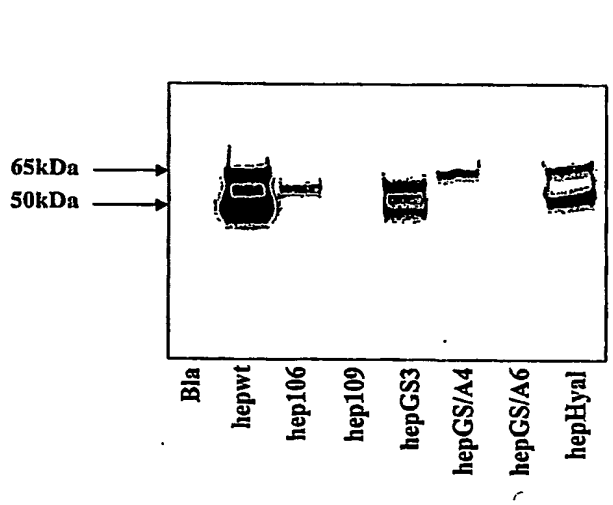

5
4
6
6
4
3
E 225
E 343
7
3
E 109
7
8
2
2
1
Q 157
8
1
Shortened loop
C-terminal domain (429-543)
Spliced segment

FIGURE 4

A)


hep106
Gln36 Pro106 Lys 158 Ile543
hep109
Gln36 Glu109 Gln157 Ile543
hepGS3
Gln36 Glu109 Gln157 Ile543
(GS)3
hepGS/A4
Gln36 Trp118 Glu143 Ile543
(GS/A)4
hepGS/A6
Gln36 Glu109 Gln157 Ile543
(GS/A)6
hepHyal
Gln36 Trp118 Glu143 Ile543
(AFKDKTP)

B)


65kDa
50kDa
Bla
hepwt
hep106
hep109
hepGS3
hepGS/A4
hepGS/A6
hepHyal

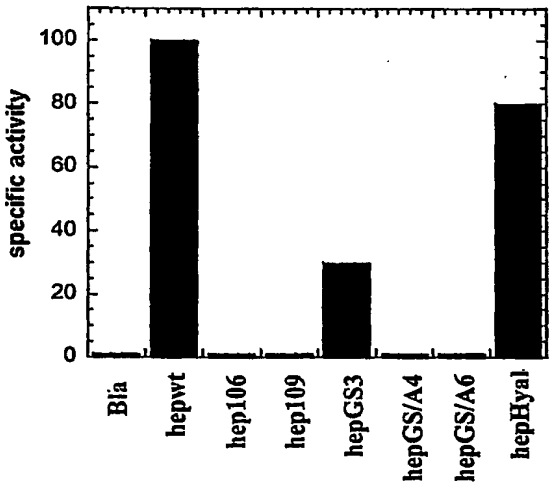

specific activity
100
80
60
40
20
0
Bla
hepwt
hep106
hep109
hepGS3
hepGS/A4
hepGS/A6
hepHyal

FIGURE 5

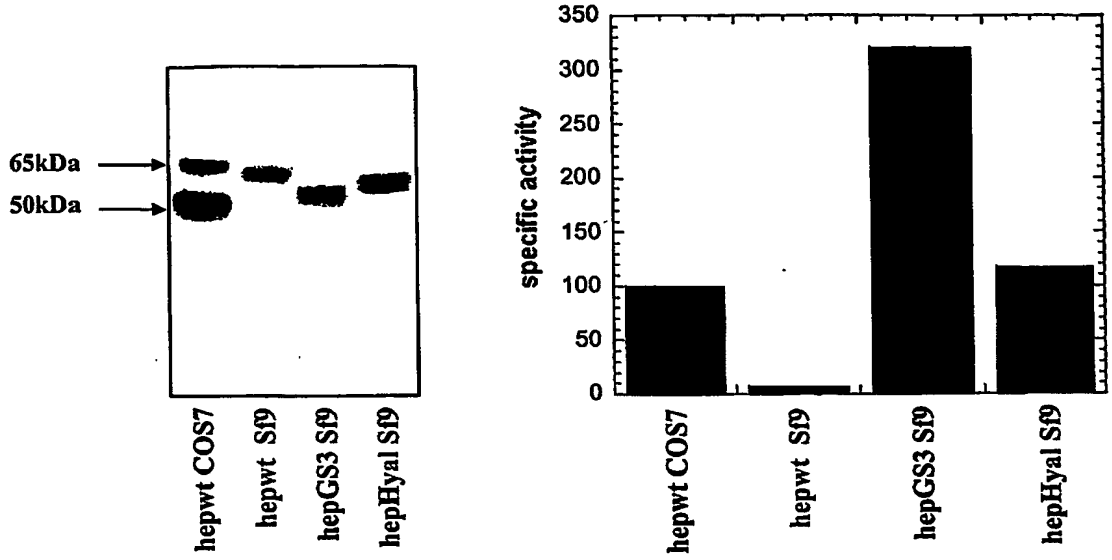

65kDa
50kDa
hepwt COS7
hepwt Sf9
hepGS3 Sf9
hepHyal Sf9
specific activity
350
300
250
200
150
100
50
0
hepwt COS7
hepwt Sf9
hepGS3 Sf9
hepHyal Sf9

FIGURE 6

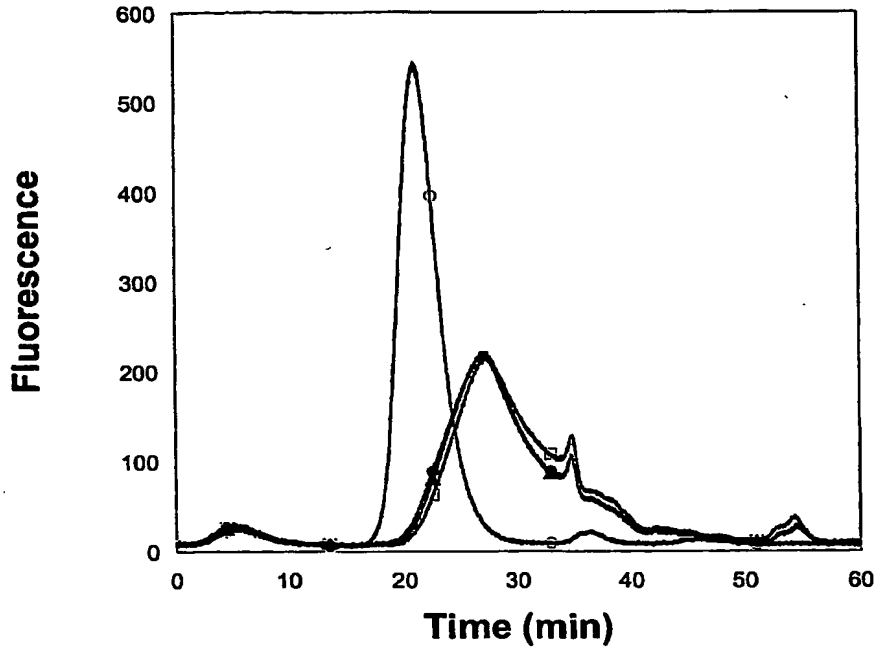

Fluorescence
600
500
400
300
200
100
0
0
10
20
30
40
50
60
Time (min)

FIGURE 7A

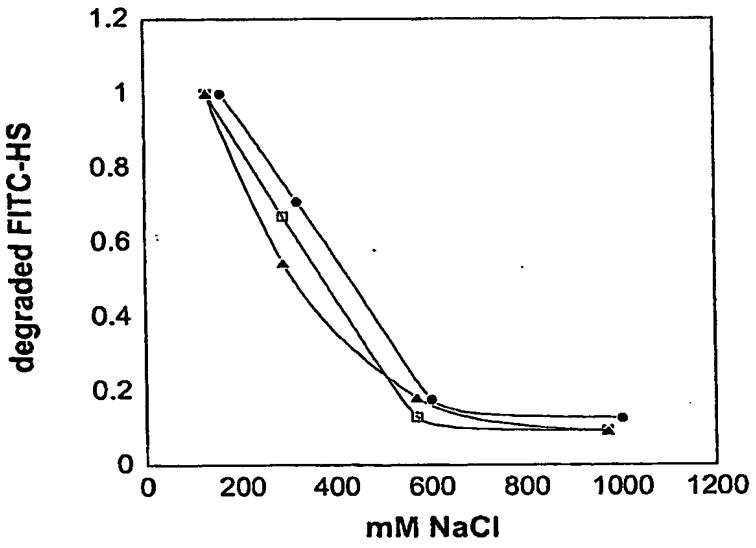

1.2
1
0.8
0.6
0.4
0.2
0
degraded FITC-HS
0
200
400
600
800
1000
1200
mM NaCl

FIGURE 7B

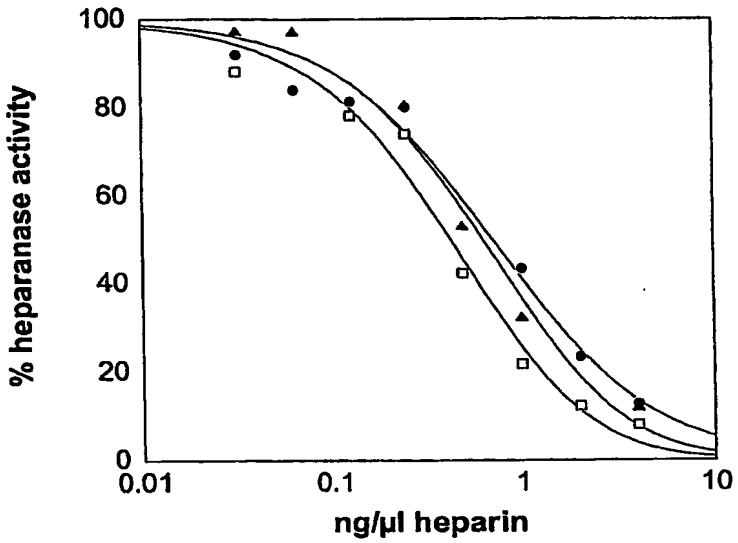
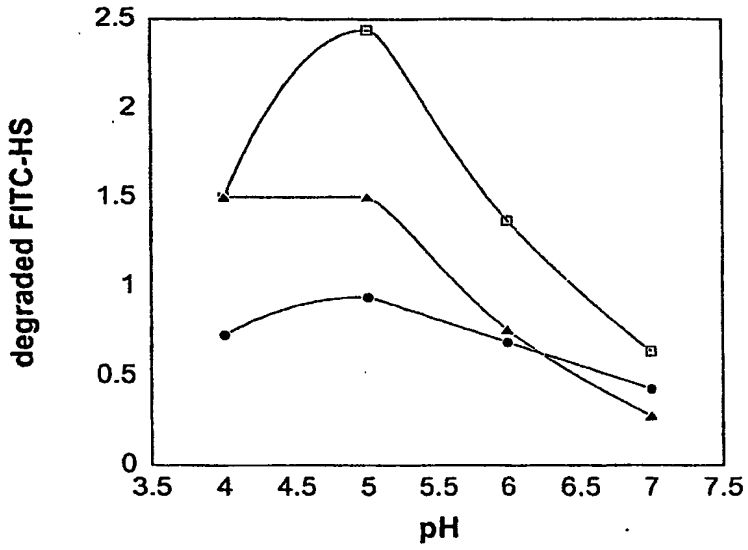


FIGURE 7C



SEQUENCE LISTING

<110> Steinkuhler, Christian
Lahm, Armin
Pallaoro, Michele
Nardella, Caterina

<120> SYNTHETIC HEPARANASE MOLECULES AND USES THEREOF

<130> ITR0060PV2

<150> 60/506,479
<151> 2003-09-26

<160> 43

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 1
cgggatccgc cgcaccatgc tgctgcgctc gaagcctgcg 40

<210> 2
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 2
tcagatgcaa gcagcaactt tggc 24

<210> 3
<211> 48
<212> DNA
<213> PCR Primer

<400> 3
ctaattttcg atcccaagaa ggaaaaaaag ttcaagaaca gcacctac 48

<210> 4
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 4
aagacagact tcctaatttt cgatcccaaa aagttcaaga acagcaccta c 51

<210> 5

<211> 57
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 5
ctaattttcg atcccaagaa ggaaggtagc ggttccggct ctaaaaagtt caagaac 57

<210> 6
<211> 87
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 6
ctaattttcg atcccaagaa ggaaggtagc ggcgctggat caggggcagc aggatccggc 60
gccaaaaagt tcaagaacag cacctac 87

<210> 7
<211> 72
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 7
acctttgaag agagaagtta ctggggttca ggggcaggat ccggcgccga atggccctac 60
caggagcaat tg 72

<210> 8
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> Peptide

<400> 8
Trp Ala Phe Lys Asp Lys Pro Thr
1 5

<210> 9
<211> 69
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 9
acctttgaag agagaagtta ctgggccttc aaggacaaga cccccgaatg gccctaccag 60
gagcaattg 69

<210> 10
<211> 15
<212> PRT

<213> Artificial Sequence

<220>
<223> Peptide

<400> 10
Glu Gly Ser Gly Ser Glu Asn Leu Tyr Phe Gln Gly Ser Gly Gly
1 5 10 15

<210> 11
<211> 63
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR Primer

<400> 11
ggcagcggat ctgagaacct gtacttccag ggttccggtt caacctttga agagagaagt 60
tac 63

<210> 12
<211> 16
<212> PRT
<213> Artificial Sequence

<220>
<223> Peptide

<400> 12
Gln Gly Ser Gly Ser Glu Asn Leu Tyr Phe Gln Gly Ser Gly Ser Lys
1 5 10 15

<210> 13
<211> 66
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR Primer

<400> 13
tctggatccg gtgaaaatct ctattttcag ggctcaggaa gtaaaaagtt caagaacagc 60
acctac 66

<210> 14
<211> 17
<212> PRT
<213> Artificial Sequence

<220>
<223> Peptide

<400> 14
Glu Pro Asn Ser Phe Leu Lys Lys Ala Asp Ile Phe Ile Asn Gly Ser
1 5 10 15
Gln

<210> 15
<211> 74
<212> PRT
<213> Human

<400> 15
Gln Asp Val Val Asp Leu Asp Phe Phe Thr Gln Glu Pro Leu His Leu
1 5 10 15
Val Ser Pro Ser Phe Leu Ser Val Thr Ile Asp Ala Asn Leu Ala Thr
20 25 30
Asp Pro Arg Phe Leu Ile Leu Leu Gly Ser Pro Lys Leu Arg Thr Leu
35 40 45
Ala Arg Gly Leu Ser Pro Ala Tyr Leu Arg Phe Gly Gly Thr Lys Thr
50 55 60
Asp Phe Leu Ile Phe Asp Pro Lys Lys Glu
65 70

<210> 16
<211> 386
<212> PRT
<213> Human

<400> 16
Lys Lys Phe Lys Asn Ser Thr Tyr Ser Arg Ser Ser Val Asp Val Leu
1 5 10 15
Tyr Thr Phe Ala Asn Cys Ser Gly Leu Asp Leu Ile Phe Gly Leu Asn
20 25 30
Ala Leu Leu Arg Thr Ala Asp Leu Gln Trp Asn Ser Ser Asn Ala Gln
35 40 45
Leu Leu Leu Asp Tyr Cys Ser Ser Lys Gly Tyr Asn Ile Ser Trp Glu
50 55 60
Leu Gly Asn Glu Pro Asn Ser Phe Leu Lys Lys Ala Asp Ile Phe Ile
65 70 75 80
Asn Gly Ser Gln Leu Gly Glu Asp Phe Ile Gln Leu His Lys Leu Leu
85 90 95
Arg Lys Ser Thr Phe Lys Asn Ala Lys Leu Tyr Gly Pro Asp Val Gly
100 105 110
Gln Pro Arg Arg Lys Thr Ala Lys Met Leu Lys Ser Phe Leu Lys Ala
115 120 125
Gly Gly Glu Val Ile Asp Ser Val Thr Trp His His Tyr Tyr Leu Asn
130 135 140
Gly Arg Thr Ala Thr Arg Glu Asp Phe Leu Asn Pro Asp Val Leu Asp
145 150 155 160
Ile Phe Ile Ser Ser Val Gln Lys Val Phe Gln Val Val Glu Ser Thr
165 170 175
Arg Pro Gly Lys Lys Val Trp Leu Gly Glu Thr Ser Ser Ala Tyr Gly
180 185 190
Gly Gly Ala Pro Leu Leu Ser Asp Thr Phe Ala Ala Gly Phe Met Trp
195 200 205
Leu Asp Lys Leu Gly Leu Ser Ala Arg Met Gly Ile Glu Val Val Met
210 215 220
Arg Gln Val Phe Phe Gly Ala Gly Asn Tyr His Leu Val Asp Glu Asn
225 230 235 240
Phe Asp Pro Leu Pro Asp Tyr Trp Leu Ser Leu Leu Phe Lys Lys Leu
245 250 255
Val Gly Thr Lys Val Leu Met Ala Ser Val Gln Gly Ser Lys Arg Arg
260 265 270
Lys Leu Arg Val Tyr Leu His Cys Thr Asn Thr Asp Asn Pro Arg Tyr
275 280 285
Lys Glu Gly Asp Leu Thr Leu Tyr Ala Ile Asn Leu His Asn Val Thr
290 295 300

Lys Tyr Leu Arg Leu Pro Tyr Pro Phe Ser Asn Lys Gln Val Asp Lys
305 310 315 320
Tyr Leu Leu Arg Pro Leu Gly Pro His Gly Leu Leu Ser Lys Ser Val
325 330 335
Gln Leu Asn Gly Leu Thr Leu Lys Met Val Asp Asp Gln Thr Leu Pro
340 345 350
Pro Leu Met Glu Lys Pro Leu Arg Pro Gly Ser Ser Leu Gly Leu Pro
355 360 365
Ala Phe Ser Tyr Ser Phe Phe Val Ile Arg Asn Ala Lys Val Ala Ala
370 375 380
Cys Ile
385

<210> 17
<211> 492
<212> PRT
<213> Artificial Sequence

<220>
<223> hep 106

<400> 17
Met Leu Leu Arg Ser Lys Pro Ala Leu Pro Pro Pro Leu Met Leu Leu
1 5 10 15
Leu Leu Gly Pro Leu Gly Pro Leu Ser Pro Gly Ala Leu Pro Arg Pro
20 25 30
Ala Gln Ala Gln Asp Val Val Asp Leu Asp Phe Phe Thr Gln Glu Pro
35 40 45
Leu His Leu Val Ser Pro Ser Phe Leu Ser Val Thr Ile Asp Ala Asn
50 55 60
Leu Ala Thr Asp Pro Arg Phe Leu Ile Leu Leu Gly Ser Pro Lys Leu
65 70 75 80
Arg Thr Leu Ala Arg Gly Leu Ser Pro Ala Tyr Leu Arg Phe Gly Gly
85 90 95
Thr Lys Thr Asp Phe Leu Ile Phe Asp Pro Lys Lys Phe Lys Asn Ser
100 105 110
Thr Tyr Ser Arg Ser Ser Val Asp Val Leu Tyr Thr Phe Ala Asn Cys
115 120 125
Ser Gly Leu Asp Leu Ile Phe Gly Leu Asn Ala Leu Leu Arg Thr Ala
130 135 140
Asp Leu Gln Trp Asn Ser Ser Asn Ala Gln Leu Leu Leu Asp Tyr Cys
145 150 155 160
Ser Ser Lys Gly Tyr Asn Ile Ser Trp Glu Leu Gly Asn Glu Pro Asn
165 170 175
Ser Phe Leu Lys Lys Ala Asp Ile Phe Ile Asn Gly Ser Gln Leu Gly
180 185 190
Glu Asp Phe Ile Gln Leu His Lys Leu Leu Arg Lys Ser Thr Phe Lys
195 200 205
Asn Ala Lys Leu Tyr Gly Pro Asp Val Gly Gln Pro Arg Arg Lys Thr
210 215 220
Ala Lys Met Leu Lys Ser Phe Leu Lys Ala Gly Gly Glu Val Ile Asp
225 230 235 240
Ser Val Thr Trp His His Tyr Tyr Leu Asn Gly Arg Thr Ala Thr Arg
245 250 255
Glu Asp Phe Leu Asn Pro Asp Val Leu Asp Ile Phe Ile Ser Ser Val
260 265 270
Gln Lys Val Phe Gln Val Val Glu Ser Thr Arg Pro Gly Lys Lys Val
275 280 285
Trp Leu Gly Glu Thr Ser Ser Ala Tyr Gly Gly Gly Ala Pro Leu Leu
290 295 300
Ser Asp Thr Phe Ala Ala Gly Phe Met Trp Leu Asp Lys Leu Gly Leu

```
305                 310                 315                 320
Ser Ala Arg Met Gly Ile Glu Val Val Met Arg Gln Val Phe Phe Gly
                325                 330                 335
Ala Gly Asn Tyr His Leu Val Asp Glu Asn Phe Asp Pro Leu Pro Asp
            340                 345                 350
Tyr Trp Leu Ser Leu Leu Phe Lys Lys Leu Val Gly Thr Lys Val Leu
        355                 360                 365
Met Ala Ser Val Gln Gly Ser Lys Arg Arg Lys Leu Arg Val Tyr Leu
    370                 375                 380
His Cys Thr Asn Thr Asp Asn Pro Arg Tyr Lys Glu Gly Asp Leu Thr
385                 390                 395                 400
Leu Tyr Ala Ile Asn Leu His Asn Val Thr Lys Tyr Leu Arg Leu Pro
                405                 410                 415
Tyr Pro Phe Ser Asn Lys Gln Val Asp Lys Tyr Leu Leu Arg Pro Leu
            420                 425                 430
Gly Pro His Gly Leu Leu Ser Lys Ser Val Gln Leu Asn Gly Leu Thr
        435                 440                 445
Leu Lys Met Val Asp Asp Gln Thr Leu Pro Pro Leu Met Glu Lys Pro
    450                 455                 460
Leu Arg Pro Gly Ser Ser Leu Gly Leu Pro Ala Phe Ser Tyr Ser Phe
465                 470                 475                 480
Phe Val Ile Arg Asn Ala Lys Val Ala Ala Cys Ile
                485                 490
```

<210> 18
<211> 1479
<212> DNA
<213> Artificial Sequence

<220>
<223> hep 106

<400> 18

```
atgctgctgc gctcgaagcc tgcgctgccg ccgccgctga tgctgctgct cctggggccg 60
ctgggtcccc tctcccctgg cgccctgccc cgacctgcgc aagcacagga cgtcgtggac 120
ctggacttct tcacccagga gccgctgcac ctggtgagcc cctcgttcct gtccgtcacc 180
attgacgcca acctggccac ggacccgcgg ttcctcatcc tcctgggttc tccaaagctt 240
cgtaccttgg ccagaggctt gtctcctgcg tacctgaggt ttggtggcac caagacagac 300
ttcctaattt tcgatcccaa aaagttcaag aacagcacct actcaagaag ctctgtagat 360
gtgctataca cttttgcaaa ctgctcagga ctggacttga tctttggcct aaatgcgtta 420
ttaagaacag cagatttgca gtggaacagt tctaatgctc agttgctcct ggactactgc 480
tcttccaagg ggtataacat ttcttgggaa ctaggcaatg aacctaacag tttccttaag 540
aaggctgata ttttcatcaa tgggtcgcag ttaggagaag attttattca attgcataaa 600
cttctaagaa agtccacctt caaaaatgca aaactctatg gtcctgatgt tggtcagcct 660
cgaagaaaga cggctaagat gctgaagagc ttcctgaagg ctggtggaga agtgattgat 720
tcagttacat ggcatcacta ctatttgaat ggacggactg ctaccaggga agattttcta 780
aaccctgatg tattggacat ttttatttca tctgtgcaaa aagttttcca ggtggttgag 840
agcaccaggc ctggcaagaa ggtctggtta ggagaaacaa gctctgcata tggaggcgga 900
gcgcccttgc tatccgacac ctttgcagct ggctttatgt ggctggataa attgggcctg 960
tcagcccgaa tgggaataga agtggtgatg aggcaagtat tctttggagc aggaaactac 1020
catttagtgg atgaaaactt cgatccttta cctgattatt ggctatctct tctgttcaag 1080
aaattggtgg gcaccaaggt gttaatggca agcgtgcaag gttcaaagag aaggaagctt 1140
cgagtatacc ttcattgcac aaacactgac aatccaaggt ataaagaagg agatttaact 1200
ctgtatgcca taaacctcca taatgtcacc aagtacttgc ggttacccta tccttttttct 1260
aacaagcaag tggataaata ccttctaaga cctttgggac ctcatggatt actttccaaa 1320
tctgtccaac tcaatggtct aactctaaag atggtggatg atcaaacctt gccaccttta 1380
atggaaaaac ctctccggcc aggaagttca ctgggcttgc cagctttctc atatagtttt 1440
tttgtgataa gaaatgccaa agttgctgct tgcatctga                        1479
```

<210> 19
<211> 1488

<212> DNA
<213> Artificial Sequence

<220>
<223> hep 109

<400> 19
atgctgctgc gctcgaagcc tgcgctgccg ccgccgctga tgctgctgct cctggggccg 60
ctgggtcccc tctcccctgg cgccctgccc cgacctgcgc aagcacagga cgtcgtggac 120
ctggacttct tcacccagga gccgctgcac ctggtgagcc cctcgttcct gtccgtcacc 180
attgacgcca acctggccac ggacccgcgg ttcctcatcc tcctgggttc tccaaagctt 240
cgtaccttgg ccagaggctt gtctcctgcg tacctgaggt ttggtggcac caagacagac 300
ttcctaattt tcgatcccaa gaaggaaaaa aagttcaaga acagcaccta ctcaagaagc 360
tctgtagatg tgctatacac ttttgcaaac tgctcaggac tggacttgat ctttggccta 420
aatgcgttat taagaacagc agatttgcag tggaacagtt ctaatgctca gttgctcctg 480
gactactgct cttccaaggg gtataacatt tcttgggaac taggcaatga acctaacagt 540
ttccttaaga aggctgatat tttcatcaat gggtcgcagt taggagaaga ttttattcaa 600
ttgcataaac ttctaagaaa gtccaccttc aaaaatgcaa aactctatgg tcctgatgtt 660
ggtcagcctc gaagaaagac ggctaagatg ctgaagagct tcctgaaggc tggtggagaa 720
gtgattgatt cagttacatg gcatcactac tatttgaatg gacggactgc taccagggaa 780
gattttctaa accctgatgt attggacatt tttatttcat ctgtgcaaaa agttttccag 840
gtggttgaga gcaccaggcc tggcaagaag gtctggttag gagaaacaag ctctgcatat 900
ggaggcggag cgcccttgct atccgacacc tttgcagctg gctttatgtg gctggataaa 960
ttgggcctgt cagcccgaat gggaatagaa gtggtgatga ggcaagtatt ctttggagca 1020
ggaaactacc atttagtgga tgaaaacttc gatcctttac ctgattattg gctatctctt 1080
ctgttcaaga aattggtggg caccaaggtg ttaatggcaa gcgtgcaagg ttcaaagaga 1140
aggaagcttc gagtatacct tcattgcaca aacactgaca atccaaggta taaagaagga 1200
gatttaactc tgtatgccat aaacctccat aatgtcacca agtacttgcg gttaccctat 1260
ccttttttcta acaagcaagt ggataaatac cttctaagac ctttgggacc tcatggatta 1320
ctttccaaat ctgtccaact caatggtcta actctaaaga tggtggatga tcaaaccttg 1380
ccacctttaa tggaaaaacc tctccggcca ggaagttcac tgggcttgcc agctttctca 1440
tatagttttt ttgtgataag aaatgccaaa gttgctgctt gcatctga 1488

<210> 20
<211> 495
<212> PRT
<213> Artificial Sequence

<220>
<223> hep 109

<400> 20
Met Leu Leu Arg Ser Lys Pro Ala Leu Pro Pro Pro Leu Met Leu Leu
1 5 10 15
Leu Leu Gly Pro Leu Gly Pro Leu Ser Pro Gly Ala Leu Pro Arg Pro
20 25 30
Ala Gln Ala Gln Asp Val Val Asp Leu Asp Phe Phe Thr Gln Glu Pro
35 40 45
Leu His Leu Val Ser Pro Ser Phe Leu Ser Val Thr Ile Asp Ala Asn
50 55 60
Leu Ala Thr Asp Pro Arg Phe Leu Ile Leu Leu Gly Ser Pro Lys Leu
65 70 75 80
Arg Thr Leu Ala Arg Gly Leu Ser Pro Ala Tyr Leu Arg Phe Gly Gly
85 90 95
Thr Lys Thr Asp Phe Leu Ile Phe Asp Pro Lys Lys Glu Lys Lys Phe
100 105 110
Lys Asn Ser Thr Tyr Ser Arg Ser Ser Val Asp Val Leu Tyr Thr Phe
115 120 125
Ala Asn Cys Ser Gly Leu Asp Leu Ile Phe Gly Leu Asn Ala Leu Leu
130 135 140
Arg Thr Ala Asp Leu Gln Trp Asn Ser Ser Asn Ala Gln Leu Leu Leu
145 150 155 160

```
Asp Tyr Cys Ser Ser Lys Gly Tyr Asn Ile Ser Trp Glu Leu Gly Asn
                165                 170                 175
Glu Pro Asn Ser Phe Leu Lys Lys Ala Asp Ile Phe Ile Asn Gly Ser
            180                 185                 190
Gln Leu Gly Glu Asp Phe Ile Gln Leu His Lys Leu Leu Arg Lys Ser
        195                 200                 205
Thr Phe Lys Asn Ala Lys Leu Tyr Gly Pro Asp Val Gly Gln Pro Arg
    210                 215                 220
Arg Lys Thr Ala Lys Met Leu Lys Ser Phe Leu Lys Ala Gly Gly Glu
225                 230                 235                 240
Val Ile Asp Ser Val Thr Trp His His Tyr Tyr Leu Asn Gly Arg Thr
                245                 250                 255
Ala Thr Arg Glu Asp Phe Leu Asn Pro Asp Val Leu Asp Ile Phe Ile
            260                 265                 270
Ser Ser Val Gln Lys Val Phe Gln Val Val Glu Ser Thr Arg Pro Gly
        275                 280                 285
Lys Lys Val Trp Leu Gly Glu Thr Ser Ser Ala Tyr Gly Gly Gly Ala
    290                 295                 300
Pro Leu Leu Ser Asp Thr Phe Ala Ala Gly Phe Met Trp Leu Asp Lys
305                 310                 315                 320
Leu Gly Leu Ser Ala Arg Met Gly Ile Glu Val Val Met Arg Gln Val
                325                 330                 335
Phe Phe Gly Ala Gly Asn Tyr His Leu Val Asp Glu Asn Phe Asp Pro
            340                 345                 350
Leu Pro Asp Tyr Trp Leu Ser Leu Leu Phe Lys Lys Leu Val Gly Thr
        355                 360                 365
Lys Val Leu Met Ala Ser Val Gln Gly Ser Lys Arg Arg Lys Leu Arg
    370                 375                 380
Val Tyr Leu His Cys Thr Asn Thr Asp Asn Pro Arg Tyr Lys Glu Gly
385                 390                 395                 400
Asp Leu Thr Leu Tyr Ala Ile Asn Leu His Asn Val Thr Lys Tyr Leu
                405                 410                 415
Arg Leu Pro Tyr Pro Phe Ser Asn Lys Gln Val Asp Lys Tyr Leu Leu
            420                 425                 430
Arg Pro Leu Gly Pro His Gly Leu Leu Ser Lys Ser Val Gln Leu Asn
        435                 440                 445
Gly Leu Thr Leu Lys Met Val Asp Asp Gln Thr Leu Pro Pro Leu Met
    450                 455                 460
Glu Lys Pro Leu Arg Pro Gly Ser Ser Leu Gly Leu Pro Ala Phe Ser
465                 470                 475                 480
Tyr Ser Phe Phe Val Ile Arg Asn Ala Lys Val Ala Ala Cys Ile
                485                 490                 495
```

<210> 21
<211> 501
<212> PRT
<213> Artificial Sequence

<220>
<223> hep GS3

```
<400> 21
Met Leu Leu Arg Ser Lys Pro Ala Leu Pro Pro Pro Leu Met Leu Leu
 1               5                  10                  15
Leu Leu Gly Pro Leu Gly Pro Leu Ser Pro Gly Ala Leu Pro Arg Pro
            20                  25                  30
Ala Gln Ala Gln Asp Val Val Asp Leu Asp Phe Phe Thr Gln Glu Pro
        35                  40                  45
Leu His Leu Val Ser Pro Ser Phe Leu Ser Val Thr Ile Asp Ala Asn
    50                  55                  60
Leu Ala Thr Asp Pro Arg Phe Leu Ile Leu Leu Gly Ser Pro Lys Leu
```

```
65                  70                  75                  80
Arg Thr Leu Ala Arg Gly Leu Ser Pro Ala Tyr Leu Arg Phe Gly Gly
                85                  90                  95
Thr Lys Thr Asp Phe Leu Ile Phe Asp Pro Lys Lys Glu Gly Ser Gly
            100                 105                 110
Ser Gly Ser Lys Lys Phe Lys Asn Ser Thr Tyr Ser Arg Ser Ser Val
        115                 120                 125
Asp Val Leu Tyr Thr Phe Ala Asn Cys Ser Gly Leu Asp Leu Ile Phe
    130                 135                 140
Gly Leu Asn Ala Leu Leu Arg Thr Ala Asp Leu Gln Trp Asn Ser Ser
145                 150                 155                 160
Asn Ala Gln Leu Leu Leu Asp Tyr Cys Ser Ser Lys Gly Tyr Asn Ile
                165                 170                 175
Ser Trp Glu Leu Gly Asn Glu Pro Asn Ser Phe Leu Lys Lys Ala Asp
            180                 185                 190
Ile Phe Ile Asn Gly Ser Gln Leu Gly Glu Asp Phe Ile Gln Leu His
        195                 200                 205
Lys Leu Leu Arg Lys Ser Thr Phe Lys Asn Ala Lys Leu Tyr Gly Pro
    210                 215                 220
Asp Val Gly Gln Pro Arg Arg Lys Thr Ala Lys Met Leu Lys Ser Phe
225                 230                 235                 240
Leu Lys Ala Gly Gly Glu Val Ile Asp Ser Val Thr Trp His His Tyr
                245                 250                 255
Tyr Leu Asn Gly Arg Thr Ala Thr Arg Glu Asp Phe Leu Asn Pro Asp
            260                 265                 270
Val Leu Asp Ile Phe Ile Ser Ser Val Gln Lys Val Phe Gln Val Val
        275                 280                 285
Glu Ser Thr Arg Pro Gly Lys Lys Val Trp Leu Gly Glu Thr Ser Ser
    290                 295                 300
Ala Tyr Gly Gly Gly Ala Pro Leu Leu Ser Asp Thr Phe Ala Ala Gly
305                 310                 315                 320
Phe Met Trp Leu Asp Lys Leu Gly Leu Ser Ala Arg Met Gly Ile Glu
                325                 330                 335
Val Val Met Arg Gln Val Phe Phe Gly Ala Gly Asn Tyr His Leu Val
            340                 345                 350
Asp Glu Asn Phe Asp Pro Leu Pro Asp Tyr Trp Leu Ser Leu Leu Phe
        355                 360                 365
Lys Lys Leu Val Gly Thr Lys Val Leu Met Ala Ser Val Gln Gly Ser
    370                 375                 380
Lys Arg Arg Lys Leu Arg Val Tyr Leu His Cys Thr Asn Thr Asp Asn
385                 390                 395                 400
Pro Arg Tyr Lys Glu Gly Asp Leu Thr Leu Tyr Ala Ile Asn Leu His
                405                 410                 415
Asn Val Thr Lys Tyr Leu Arg Leu Pro Tyr Pro Phe Ser Asn Lys Gln
            420                 425                 430
Val Asp Lys Tyr Leu Leu Arg Pro Leu Gly Pro His Gly Leu Leu Ser
        435                 440                 445
Lys Ser Val Gln Leu Asn Gly Leu Thr Leu Lys Met Val Asp Asp Gln
    450                 455                 460
Thr Leu Pro Pro Leu Met Glu Lys Pro Leu Arg Pro Gly Ser Ser Leu
465                 470                 475                 480
Gly Leu Pro Ala Phe Ser Tyr Ser Phe Phe Val Ile Arg Asn Ala Lys
                485                 490                 495
Val Ala Ala Cys Ile
            500
```

<210> 22
<211> 1506
<212> DNA
<213> Artificial Sequence

<220>
<223> hep GS3

<400> 22
atgctgctgc gctcgaagcc tgcgctgccg ccgccgctga tgctgctgct cctggggccg 60
ctgggtcccc tctcccctgg cgccctgccc cgacctgcgc aagcacagga cgtcgtggac 120
ctggacttct tcacccagga gccgctgcac ctggtgagcc cctcgttcct gtccgtcacc 180
attgacgcca acctggccac ggacccgcgg ttcctcatcc tcctgggttc tccaaagctt 240
cgtaccttgg ccagaggctt gtctcctgcg tacctgaggt ttggtggcac caagacagac 300
ttcctaattt tcgatcccaa gaaggaaggt agcggttccg gctctaaaaa gttcaagaac 360
agcacctact caagaagctc tgtagatgtg ctatacactt ttgcaaactg ctcaggactg 420
gacttgatct ttggcctaaa tgcgttatta agaacagcag atttgcagtg gaacagttct 480
aatgctcagt tgctcctgga ctactgctct tccaaggggt ataacatttc ttgggaacta 540
ggcaatgaac ctaacagttt ccttaagaag gctgatattt tcatcaatgg gtcgcagtta 600
ggagaagatt ttattcaatt gcataaactt ctaagaaagt ccaccttcaa aaatgcaaaa 660
ctctatggtc ctgatgttgg tcagcctcga agaaagacgg ctaagatgct gaagagcttc 720
ctgaaggctg gtggagaagt gattgattca gttacatggc atcactacta tttgaatgga 780
cggactgcta ccagggaaga ttttctaaac cctgatgtat tggacatttt tatttcatct 840
gtgcaaaaag ttttccaggt ggttgagagc accaggcctg gcaagaaggt ctggttagga 900
gaaacaagct ctgcatatgg aggcggagcg cccttgctat ccgacacctt tgcagctggc 960
tttatgtggc tggataaatt gggcctgtca gcccgaatgg gaatagaagt ggtgatgagg 1020
caagtattct ttggagcagg aaactaccat ttagtggatg aaaacttcga tcctttacct 1080
gattattggc tatctcttct gttcaagaaa ttggtgggca ccaaggtgtt aatggcaagc 1140
gtgcaaggtt caaagagaag gaagcttcga gtataccttc attgcacaaa cactgacaat 1200
ccaaggtata aagaaggaga tttaactctg tatgccataa acctccataa tgtcaccaag 1260
tacttgcggt taccctatcc tttttctaac aagcaagtgg ataaatacct tctaagacct 1320
ttgggacctc atggattact ttccaaatct gtccaactca atggtctaac tctaaagatg 1380
gtggatgatc aaaccttgcc acctttaatg gaaaaacctc tccggccagg aagttcactg 1440
ggcttgccag ctttctcata tagtttttt gtgataagaa atgccaaagt tgctgcttgc 1500
atctga 1506

<210> 23
<211> 1584
<212> DNA
<213> Artificial Sequence

<220>
<223> hep GS4

<400> 23
atgctgctgc gctcgaagcc tgcgctgccg ccgccgctga tgctgctgct cctggggccg 60
ctgggtcccc tctcccctgg cgccctgccc cgacctgcgc aagcacagga cgtcgtggac 120
ctggacttct tcacccagga gccgctgcac ctggtgagcc cctcgttcct gtccgtcacc 180
attgacgcca acctggccac ggacccgcgg ttcctcatcc tcctgggttc tccaaagctt 240
cgtaccttgg ccagaggctt gtctcctgcg tacctgaggt ttggtggcac caagacagac 300
ttcctaattt tcgatcccaa gaaggaatca acctttgaag agagaagtta ctggggttca 360
ggggcaggat ccggcgccga atggccctac caggagcaat tgctactccg agaacactac 420
cagaaaaagt tcaagaacag cacctactca agaagctctg tagatgtgct atacactttt 480
gcaaactgct caggactgga cttgatcttt ggcctaaatg cgttattaag aacagcagat 540
ttgcagtgga acagttctaa tgctcagttg ctcctggact actgctcttc caaggggtat 600
aacatttctt gggaactagg caatgaacct aacagtttcc ttaagaaggc tgatattttc 660
atcaatgggt cgcagttagg agaagatttt attcaattgc ataaacttct aagaaagtcc 720
accttcaaaa atgcaaaact ctatggtcct gatgttggtc agcctcgaag aaagacggct 780
aagatgctga agagcttcct gaaggctggt ggagaagtga ttgattcagt tacatggcat 840
cactactatt tgaatggacg gactgctacc agggaagatt ttctaaaccc tgatgtattg 900
gacattttta tttcatctgt gcaaaaagtt ttccaggtgg ttgagagcac caggcctggc 960
aagaaggtct ggttaggaga aacaagctct gcatatggag gcggagcgcc cttgctatcc 1020
gacacctttg cagctggctt tatgtggctg gataaattgg gcctgtcagc ccgaatggga 1080
atagaagtgg tgatgaggca agtattcttt ggagcaggaa actaccattt agtggatgaa 1140
aacttcgatc ctttacctga ttattggcta tctcttctgt tcaagaaatt ggtgggcacc 1200
aaggtgttaa tggcaagcgt gcaaggttca aagagaagga agcttcgagt ataccttcat 1260
tgcacaaaca ctgacaatcc aaggtataaa gaaggagatt taactctgta tgccataaac 1320

ctccataatg tcaccaagta cttgcggtta ccctatcctt tttctaacaa gcaagtggat 1380
aaataccttc taagaccttt gggacctcat ggattacttt ccaaatctgt ccaactcaat 1440
ggtctaactc taaagatggt ggatgatcaa accttgccac ctttaatgga aaaacctctc 1500
cggccaggaa gttcactggg cttgccagct ttctcatata gttttttttgt gataagaaat 1560
gccaaagttg ctgcttgcat ctga 1584

<210> 24
<211> 1524
<212> DNA
<213> Artificial Sequence

<220>
<223> hep GS6

<400> 24
atgctgctgc gctcgaagcc tgcgctgccg ccgccgctga tgctgctgct cctggggccg 60
ctgggtcccc tctcccctgg cgccctgccc cgacctgcgc aagcacagga cgtcgtggac 120
ctggacttct tcacccagga gccgctgcac ctggtgagcc cctcgttcct gtccgtcacc 180
attgacgcca acctggccac ggacccgcgg ttcctcatcc tcctgggttc tccaaagctt 240
cgtaccttgg ccagaggctt gtctcctgcg tacctgaggt ttggtggcac caagacagac 300
ttcctaattt tcgatcccaa gaaggaaggt agcggttccg gctctggtag cggctctggt 360
agcaaaaagt tcaagaacag cacctactca agaagctctg tagatgtgct atacactttt 420
gcaaactgct caggactgga cttgatcttt ggcctaaatg cgttattaag aacagcagat 480
ttgcagtgga acagttctaa tgctcagttg ctcctggact actgctcttc caaggggtat 540
aacatttctt gggaactagg caatgaacct aacagtttcc ttaagaaggc tgatattttc 600
atcaatgggt cgcagttagg agaagatttt attcaattgc ataaacttct aagaaagtcc 660
accttcaaaa atgcaaaact ctatggtcct gatgttggtc agcctcgaag aaagacggct 720
aagatgctga agagcttcct gaaggctggt ggagaagtga ttgattcagt tacatggcat 780
cactactatt tgaatggacg gactgctacc agggaagatt ttctaaaccc tgatgtattg 840
gacattttta tttcatctgt gcaaaaagtt ttccaggtgg ttgagagcac caggcctggc 900
aagaaggtct ggttaggaga aacaagctct gcatatggag gcggagcgcc cttgctatcc 960
gacacctttg cagctggctt tatgtggctg gataaattgg gcctgtcagc ccgaatggga 1020
atagaagtgg tgatgaggca agtattcttt ggagcaggaa actaccattt agtggatgaa 1080
aacttcgatc ctttacctga ttattggcta tctcttctgt tcaagaaatt ggtgggcacc 1140
aaggtgttaa tggcaagcgt gcaaggttca aagagaagga agcttcgagt ataccttcat 1200
tgcacaaaca ctgacaatcc aaggtataaa gaaggagatt taactctgta tgccataaac 1260
ctccataatg tcaccaagta cttgcggtta ccctatcctt tttctaacaa gcaagtggat 1320
aaataccttc taagaccttt gggacctcat ggattacttt ccaaatctgt ccaactcaat 1380
ggtctaactc taaagatggt ggatgatcaa accttgccac ctttaatgga aaaacctctc 1440
cggccaggaa gttcactggg cttgccagct ttctcatata gttttttttgt gataagaaat 1500
gccaaagttg ctgcttgcat ctga 1524

<210> 25
<211> 527
<212> PRT
<213> Artificial Sequence

<220>
<223> hep GS-A4

<400> 25
Met Leu Leu Arg Ser Lys Pro Ala Leu Pro Pro Pro Leu Met Leu Leu
 1   5   10   15
Leu Leu Gly Pro Leu Gly Pro Leu Ser Pro Gly Ala Leu Pro Arg Pro
            20   25   30
Ala Gln Ala Gln Asp Val Val Asp Leu Asp Phe Phe Thr Gln Glu Pro
        35   40   45
Leu His Leu Val Ser Pro Ser Phe Leu Ser Val Thr Ile Asp Ala Asn
    50   55   60
Leu Ala Thr Asp Pro Arg Phe Leu Ile Leu Leu Gly Ser Pro Lys Leu
65   70   75   80
Arg Thr Leu Ala Arg Gly Leu Ser Pro Ala Tyr Leu Arg Phe Gly Gly

```
                85                  90                  95
Thr Lys Thr Asp Phe Leu Ile Phe Asp Pro Lys Lys Glu Ser Thr Phe
            100                 105                 110
Glu Glu Arg Ser Tyr Trp Gly Ser Gly Ala Gly Ser Gly Ala Glu Trp
        115                 120                 125
Pro Tyr Gln Glu Gln Leu Leu Leu Arg Glu His Tyr Gln Lys Lys Phe
    130                 135                 140
Lys Asn Ser Thr Tyr Ser Arg Ser Ser Val Asp Val Leu Tyr Thr Phe
145                 150                 155                 160
Ala Asn Cys Ser Gly Leu Asp Leu Ile Phe Gly Leu Asn Ala Leu Leu
                165                 170                 175
Arg Thr Ala Asp Leu Gln Trp Asn Ser Ser Asn Ala Gln Leu Leu Leu
            180                 185                 190
Asp Tyr Cys Ser Ser Lys Gly Tyr Asn Ile Ser Trp Glu Leu Gly Asn
        195                 200                 205
Glu Pro Asn Ser Phe Leu Lys Lys Ala Asp Ile Phe Ile Asn Gly Ser
    210                 215                 220
Gln Leu Gly Glu Asp Phe Ile Gln Leu His Lys Leu Leu Arg Lys Ser
225                 230                 235                 240
Thr Phe Lys Asn Ala Lys Leu Tyr Gly Pro Asp Val Gly Gln Pro Arg
                245                 250                 255
Arg Lys Thr Ala Lys Met Leu Lys Ser Phe Leu Lys Ala Gly Gly Glu
            260                 265                 270
Val Ile Asp Ser Val Thr Trp His His Tyr Tyr Leu Asn Gly Arg Thr
        275                 280                 285
Ala Thr Arg Glu Asp Phe Leu Asn Pro Asp Val Leu Asp Ile Phe Ile
    290                 295                 300
Ser Ser Val Gln Lys Val Phe Gln Val Val Glu Ser Thr Arg Pro Gly
305                 310                 315                 320
Lys Lys Val Trp Leu Gly Glu Thr Ser Ser Ala Tyr Gly Gly Gly Ala
                325                 330                 335
Pro Leu Leu Ser Asp Thr Phe Ala Ala Gly Phe Met Trp Leu Asp Lys
            340                 345                 350
Leu Gly Leu Ser Ala Arg Met Gly Ile Glu Val Val Met Arg Gln Val
        355                 360                 365
Phe Phe Gly Ala Gly Asn Tyr His Leu Val Asp Glu Asn Phe Asp Pro
    370                 375                 380
Leu Pro Asp Tyr Trp Leu Ser Leu Leu Phe Lys Lys Leu Val Gly Thr
385                 390                 395                 400
Lys Val Leu Met Ala Ser Val Gln Gly Ser Lys Arg Arg Lys Leu Arg
                405                 410                 415
Val Tyr Leu His Cys Thr Asn Thr Asp Asn Pro Arg Tyr Lys Glu Gly
            420                 425                 430
Asp Leu Thr Leu Tyr Ala Ile Asn Leu His Asn Val Thr Lys Tyr Leu
        435                 440                 445
Arg Leu Pro Tyr Pro Phe Ser Asn Lys Gln Val Asp Lys Tyr Leu Leu
    450                 455                 460
Arg Pro Leu Gly Pro His Gly Leu Leu Ser Lys Ser Val Gln Leu Asn
465                 470                 475                 480
Gly Leu Thr Leu Lys Met Val Asp Asp Gln Thr Leu Pro Pro Leu Met
                485                 490                 495
Glu Lys Pro Leu Arg Pro Gly Ser Ser Leu Gly Leu Pro Ala Phe Ser
            500                 505                 510
Tyr Ser Phe Phe Val Ile Arg Asn Ala Lys Val Ala Ala Cys Ile
        515                 520                 525
```

<210> 26
<211> 507
<212> PRT
<213> Artificial Sequence

<220>
<223> hep GS-A6

<400> 26

```
Met Leu Leu Arg Ser Lys Pro Ala Leu Pro Pro Pro Leu Met Leu Leu
 1               5                  10                  15
Leu Leu Gly Pro Leu Gly Pro Leu Ser Pro Gly Ala Leu Pro Arg Pro
            20                  25                  30
Ala Gln Ala Gln Asp Val Val Asp Leu Asp Phe Phe Thr Gln Glu Pro
        35                  40                  45
Leu His Leu Val Ser Pro Ser Phe Leu Ser Val Thr Ile Asp Ala Asn
    50                  55                  60
Leu Ala Thr Asp Pro Arg Phe Leu Ile Leu Leu Gly Ser Pro Lys Leu
65                  70                  75                  80
Arg Thr Leu Ala Arg Gly Leu Ser Pro Ala Tyr Leu Arg Phe Gly Gly
                85                  90                  95
Thr Lys Thr Asp Phe Leu Ile Phe Asp Pro Lys Lys Glu Gly Ser Gly
            100                 105                 110
Ser Gly Ser Gly Ser Gly Ser Gly Ser Lys Lys Phe Lys Asn Ser Thr
        115                 120                 125
Tyr Ser Arg Ser Ser Val Asp Val Leu Tyr Thr Phe Ala Asn Cys Ser
    130                 135                 140
Gly Leu Asp Leu Ile Phe Gly Leu Asn Ala Leu Leu Arg Thr Ala Asp
145                 150                 155                 160
Leu Gln Trp Asn Ser Ser Asn Ala Gln Leu Leu Leu Asp Tyr Cys Ser
                165                 170                 175
Ser Lys Gly Tyr Asn Ile Ser Trp Glu Leu Gly Asn Glu Pro Asn Ser
            180                 185                 190
Phe Leu Lys Lys Ala Asp Ile Phe Ile Asn Gly Ser Gln Leu Gly Glu
        195                 200                 205
Asp Phe Ile Gln Leu His Lys Leu Leu Arg Lys Ser Thr Phe Lys Asn
    210                 215                 220
Ala Lys Leu Tyr Gly Pro Asp Val Gly Gln Pro Arg Arg Lys Thr Ala
225                 230                 235                 240
Lys Met Leu Lys Ser Phe Leu Lys Ala Gly Gly Glu Val Ile Asp Ser
                245                 250                 255
Val Thr Trp His His Tyr Tyr Leu Asn Gly Arg Thr Ala Thr Arg Glu
            260                 265                 270
Asp Phe Leu Asn Pro Asp Val Leu Asp Ile Phe Ile Ser Ser Val Gln
        275                 280                 285
Lys Val Phe Gln Val Val Glu Ser Thr Arg Pro Gly Lys Lys Val Trp
    290                 295                 300
Leu Gly Glu Thr Ser Ser Ala Tyr Gly Gly Gly Ala Pro Leu Leu Ser
305                 310                 315                 320
Asp Thr Phe Ala Ala Gly Phe Met Trp Leu Asp Lys Leu Gly Leu Ser
                325                 330                 335
Ala Arg Met Gly Ile Glu Val Val Met Arg Gln Val Phe Phe Gly Ala
            340                 345                 350
Gly Asn Tyr His Leu Val Asp Glu Asn Phe Asp Pro Leu Pro Asp Tyr
        355                 360                 365
Trp Leu Ser Leu Leu Phe Lys Lys Leu Val Gly Thr Lys Val Leu Met
    370                 375                 380
Ala Ser Val Gln Gly Ser Lys Arg Arg Lys Leu Arg Val Tyr Leu His
385                 390                 395                 400
Cys Thr Asn Thr Asp Asn Pro Arg Tyr Lys Glu Gly Asp Leu Thr Leu
                405                 410                 415
Tyr Ala Ile Asn Leu His Asn Val Thr Lys Tyr Leu Arg Leu Pro Tyr
            420                 425                 430
Pro Phe Ser Asn Lys Gln Val Asp Lys Tyr Leu Leu Arg Pro Leu Gly
        435                 440                 445
Pro His Gly Leu Leu Ser Lys Ser Val Gln Leu Asn Gly Leu Thr Leu
    450                 455                 460
```

```
Lys Met Val Asp Asp Gln Thr Leu Pro Pro Leu Met Glu Lys Pro Leu
465                 470                 475                 480
Arg Pro Gly Ser Ser Leu Gly Leu Pro Ala Phe Ser Tyr Ser Phe Phe
                485                 490                 495
Val Ile Arg Asn Ala Lys Val Ala Ala Cys Ile
            500                 505
```

<210> 27
<211> 526
<212> PRT
<213> Artificial Sequence

<220>
<223> hep Hyal

<400> 27

```
Met Leu Leu Arg Ser Lys Pro Ala Leu Pro Pro Pro Leu Met Leu Leu
 1               5                  10                  15
Leu Leu Gly Pro Leu Gly Pro Leu Ser Pro Gly Ala Leu Pro Arg Pro
            20                  25                  30
Ala Gln Ala Gln Asp Val Val Asp Leu Asp Phe Phe Thr Gln Glu Pro
        35                  40                  45
Leu His Leu Val Ser Pro Ser Phe Leu Ser Val Thr Ile Asp Ala Asn
    50                  55                  60
Leu Ala Thr Asp Pro Arg Phe Leu Ile Leu Leu Gly Ser Pro Lys Leu
65                  70                  75                  80
Arg Thr Leu Ala Arg Gly Leu Ser Pro Ala Tyr Leu Arg Phe Gly Gly
                85                  90                  95
Thr Lys Thr Asp Phe Leu Ile Phe Asp Pro Lys Lys Glu Ser Thr Phe
            100                 105                 110
Glu Glu Arg Ser Tyr Trp Ala Phe Lys Asp Lys Thr Pro Glu Trp Pro
        115                 120                 125
Tyr Gln Glu Gln Leu Leu Leu Arg Glu His Tyr Gln Lys Lys Phe Lys
    130                 135                 140
Asn Ser Thr Tyr Ser Arg Ser Ser Val Asp Val Leu Tyr Thr Phe Ala
145                 150                 155                 160
Asn Cys Ser Gly Leu Asp Leu Ile Phe Gly Leu Asn Ala Leu Leu Arg
                165                 170                 175
Thr Ala Asp Leu Gln Trp Asn Ser Ser Asn Ala Gln Leu Leu Leu Asp
            180                 185                 190
Tyr Cys Ser Ser Lys Gly Tyr Asn Ile Ser Trp Glu Leu Gly Asn Glu
        195                 200                 205
Pro Asn Ser Phe Leu Lys Lys Ala Asp Ile Phe Ile Asn Gly Ser Gln
    210                 215                 220
Leu Gly Glu Asp Phe Ile Gln Leu His Lys Leu Leu Arg Lys Ser Thr
225                 230                 235                 240
Phe Lys Asn Ala Lys Leu Tyr Gly Pro Asp Val Gly Gln Pro Arg Arg
                245                 250                 255
Lys Thr Ala Lys Met Leu Lys Ser Phe Leu Lys Ala Gly Gly Glu Val
            260                 265                 270
Ile Asp Ser Val Thr Trp His His Tyr Tyr Leu Asn Gly Arg Thr Ala
        275                 280                 285
Thr Arg Glu Asp Phe Leu Asn Pro Asp Val Leu Asp Ile Phe Ile Ser
    290                 295                 300
Ser Val Gln Lys Val Phe Gln Val Val Glu Ser Thr Arg Pro Gly Lys
305                 310                 315                 320
Lys Val Trp Leu Gly Glu Thr Ser Ser Ala Tyr Gly Gly Gly Ala Pro
                325                 330                 335
Leu Leu Ser Asp Thr Phe Ala Ala Gly Phe Met Trp Leu Asp Lys Leu
            340                 345                 350
Gly Leu Ser Ala Arg Met Gly Ile Glu Val Val Met Arg Gln Val Phe
```

```
        355                 360                 365
Phe Gly Ala Gly Asn Tyr His Leu Val Asp Glu Asn Phe Asp Pro Leu
    370                 375                 380
Pro Asp Tyr Trp Leu Ser Leu Leu Phe Lys Lys Leu Val Gly Thr Lys
385                 390                 395                 400
Val Leu Met Ala Ser Val Gln Gly Ser Lys Arg Arg Lys Leu Arg Val
                405                 410                 415
Tyr Leu His Cys Thr Asn Thr Asp Asn Pro Arg Tyr Lys Glu Gly Asp
            420                 425                 430
Leu Thr Leu Tyr Ala Ile Asn Leu His Asn Val Thr Lys Tyr Leu Arg
        435                 440                 445
Leu Pro Tyr Pro Phe Ser Asn Lys Gln Val Asp Lys Tyr Leu Leu Arg
    450                 455                 460
Pro Leu Gly Pro His Gly Leu Leu Ser Lys Ser Val Gln Leu Asn Gly
465                 470                 475                 480
Leu Thr Leu Lys Met Val Asp Asp Gln Thr Leu Pro Pro Leu Met Glu
                485                 490                 495
Lys Pro Leu Arg Pro Gly Ser Ser Leu Gly Leu Pro Ala Phe Ser Tyr
            500                 505                 510
Ser Phe Phe Val Ile Arg Asn Ala Lys Val Ala Ala Cys Ile
        515                 520                 525
```

<210> 28
<211> 1581
<212> DNA
<213> Artificial Sequence

<220>
<223> hep Hyal

<400> 28

```
atgctgctgc gctcgaagcc tgcgctgccg ccgccgctga tgctgctgct cctggggccg 60
ctgggtcccc tctcccctgg cgccctgccc cgacctgcgc aagcacagga cgtcgtggac 120
ctggacttct tcacccagga gccgctgcac ctggtgagcc cctcgttcct gtccgtcacc 180
attgacgcca acctggccac ggacccgcgg ttcctcatcc tcctgggttc tccaaagctt 240
cgtaccttgg ccagaggctt gtctcctgcg tacctgaggt ttggtggcac caagacagac 300
ttcctaattt tcgatcccaa gaaggaatca acctttgaag agagaagtta ctgggccttc 360
aaggacaaga cccccgaatg gccctaccag gagcaattgc tactccgaga acactaccag 420
aaaaagttca agaacagcac ctactcaaga agctctgtag atgtgctata cacttttgca 480
aactgctcag gactggactt gatctttggc ctaaatgcgt tattaagaac agcagatttg 540
cagtggaaca gttctaatgc tcagttgctc ctggactact gctcttccaa ggggtataac 600
atttcttggg aactaggcaa tgaacctaac agtttcctta agaaggctga tattttcatc 660
aatgggtcgc agttaggaga agattttatt caattgcata aacttctaag aaagtccacc 720
ttcaaaaatg caaaactcta tggtcctgat gttggtcagc ctcgaagaaa gacggctaag 780
atgctgaaga gcttcctgaa ggctggtgga gaagtgattg attcagttac atggcatcac 840
tactatttga atggacggac tgctaccagg gaagattttc taaaccctga tgtattggac 900
atttttattt catctgtgca aaaagttttc caggtggttg agagcaccag gcctggcaag 960
aaggtctggt taggagaaac aagctctgca tatggaggcg gagcgccctt gctatccgac 1020
acctttgcag ctggctttat gtggctggat aaattgggcc tgtcagcccg aatgggaata 1080
gaagtggtga tgaggcaagt attctttgga gcaggaaact accatttagt ggatgaaaac 1140
ttcgatcctt tacctgatta ttggctatct cttctgttca agaaattggt gggcaccaag 1200
gtgttaatgg caagcgtgca aggttcaaag agaaggaagc ttcgagtata ccttcattgc 1260
acaaacactg acaatccaag gtataaagaa ggagatttaa ctctgtatgc cataaacctc 1320
cataatgtca ccaagtactt gcggttaccc tatccttttt ctaacaagca agtggataaa 1380
taccttctaa gacctttggg acctcatgga ttactttcca aatctgtcca actcaatggt 1440
ctaactctaa agatggtgga tgatcaaacc ttgccacctt taatggaaaa acctctccgg 1500
ccaggaagtt cactgggctt gccagctttc tcatatagtt tttttgtgat aagaaatgcc 1560
aaagttgctg cttgcatctg a                                          1581
```

<210> 29
<211> 570

<212> PRT
<213> Artificial Sequence

<220>
<223> hep TEV110-158

<400> 29

```
Met Leu Leu Arg Ser Lys Pro Ala Leu Pro Pro Pro Leu Met Leu Leu
 1               5                  10                  15
Leu Leu Gly Pro Leu Gly Pro Leu Ser Pro Gly Ala Leu Pro Arg Pro
            20                  25                  30
Ala Gln Ala Gln Asp Val Val Asp Leu Asp Phe Phe Thr Gln Glu Pro
        35                  40                  45
Leu His Leu Val Ser Pro Ser Phe Leu Ser Val Thr Ile Asp Ala Asn
    50                  55                  60
Leu Ala Thr Asp Pro Arg Phe Leu Ile Leu Leu Gly Ser Pro Lys Leu
65                  70                  75                  80
Arg Thr Leu Ala Arg Gly Leu Ser Pro Ala Tyr Leu Arg Phe Gly Gly
                85                  90                  95
Thr Lys Thr Asp Phe Leu Ile Phe Asp Pro Lys Lys Glu Gly Ser Gly
            100                 105                 110
Ser Glu Asn Leu Tyr Phe Gln Gly Ser Gly Ser Thr Phe Glu Glu Arg
        115                 120                 125
Ser Tyr Trp Gln Ser Gln Val Asn Gln Asp Ile Cys Lys Tyr Gly Ser
    130                 135                 140
Ile Pro Pro Asp Val Glu Glu Lys Leu Arg Leu Glu Trp Pro Tyr Gln
145                 150                 155                 160
Glu Gln Leu Leu Leu Arg Glu His Tyr Gln Ser Gly Ser Gly Glu Asn
                165                 170                 175
Leu Tyr Phe Gln Gly Ser Gly Ser Lys Lys Phe Lys Asn Ser Thr Tyr
            180                 185                 190
Ser Arg Ser Ser Val Asp Val Leu Tyr Thr Phe Ala Asn Cys Ser Gly
        195                 200                 205
Leu Asp Leu Ile Phe Gly Leu Asn Ala Leu Leu Arg Thr Ala Asp Leu
    210                 215                 220
Gln Trp Asn Ser Ser Asn Ala Gln Leu Leu Leu Asp Tyr Cys Ser Ser
225                 230                 235                 240
Lys Gly Tyr Asn Ile Ser Trp Glu Leu Gly Asn Glu Pro Asn Ser Phe
                245                 250                 255
Leu Lys Lys Ala Asp Ile Phe Ile Asn Gly Ser Gln Leu Gly Glu Asp
            260                 265                 270
Phe Ile Gln Leu His Lys Leu Leu Arg Lys Ser Thr Phe Lys Asn Ala
        275                 280                 285
Lys Leu Tyr Gly Pro Asp Val Gly Gln Pro Arg Arg Lys Thr Ala Lys
    290                 295                 300
Met Leu Lys Ser Phe Leu Lys Ala Gly Gly Glu Val Ile Asp Ser Val
305                 310                 315                 320
Thr Trp His His Tyr Tyr Leu Asn Gly Arg Thr Ala Thr Arg Glu Asp
                325                 330                 335
Phe Leu Asn Pro Asp Val Leu Asp Ile Phe Ile Ser Ser Val Gln Lys
            340                 345                 350
Val Phe Gln Val Val Glu Ser Thr Arg Pro Gly Lys Lys Val Trp Leu
        355                 360                 365
Gly Glu Thr Ser Ser Ala Tyr Gly Gly Gly Ala Pro Leu Leu Ser Asp
    370                 375                 380
Thr Phe Ala Ala Gly Phe Met Trp Leu Asp Lys Leu Gly Leu Ser Ala
385                 390                 395                 400
Arg Met Gly Ile Glu Val Val Met Arg Gln Val Phe Phe Gly Ala Gly
                405                 410                 415
Asn Tyr His Leu Val Asp Glu Asn Phe Asp Pro Leu Pro Asp Tyr Trp
            420                 425                 430
Leu Ser Leu Leu Phe Lys Lys Leu Val Gly Thr Lys Val Leu Met Ala
```

```
        435                 440                 445
Ser Val Gln Gly Ser Lys Arg Arg Lys Leu Arg Val Tyr Leu His Cys
    450                 455                 460
Thr Asn Thr Asp Asn Pro Arg Tyr Lys Glu Gly Asp Leu Thr Leu Tyr
465                 470                 475                 480
Ala Ile Asn Leu His Asn Val Thr Lys Tyr Leu Arg Leu Pro Tyr Pro
                485                 490                 495
Phe Ser Asn Lys Gln Val Asp Lys Tyr Leu Leu Arg Pro Leu Gly Pro
            500                 505                 510
His Gly Leu Leu Ser Lys Ser Val Gln Leu Asn Gly Leu Thr Leu Lys
        515                 520                 525
Met Val Asp Asp Gln Thr Leu Pro Pro Leu Met Glu Lys Pro Leu Arg
    530                 535                 540
Pro Gly Ser Ser Leu Gly Leu Pro Ala Phe Ser Tyr Ser Phe Phe Val
545                 550                 555                 560
Ile Arg Asn Ala Lys Val Ala Ala Cys Ile
                565                 570
```

<210> 30
<211> 1668
<212> DNA
<213> Artificial Sequence

<220>
<223> hep TEV110

<400> 30

```
atgctgctgc gctcgaagcc tgcgctgccg ccgccgctga tgctgctgct cctggggccg 60
ctgggtcccc tctcccctgg cgccctgccc cgacctgcgc aagcacagga cgtcgtggac 120
ctggacttct tcacccagga gccgctgcac ctggtgagcc cctcgttcct gtccgtcacc 180
attgacgcca acctggccac ggacccgcgg ttcctcatcc tcctgggttc tccaaagctt 240
cgtaccttgg ccagaggctt gtctcctgcg tacctgaggt ttggtggcac caagacagac 300
ttcctaattt tcgatcccaa gaaggaaggc agcggatctg agaacctgta cttccagggt 360
tccggttcaa cctttgaaga gagaagttac tggcaatctc aagtcaacca ggatatttgc 420
aaatatggat ccatccctcc tgatgtggag gagaagttac ggttggaatg gccctaccag 480
gagcaattgc tactccgaga acactaccag aaaaagttca agaacagcac ctactcaaga 540
agctctgtag atgtgctata cacttttgca aactgctcag gactggactt gatctttggc 600
ctaaatgcgt tattaagaac agcagatttg cagtggaaca gttctaatgc tcagttgctc 660
ctggactact gctcttccaa ggggtataac atttcttggg aactaggcaa tgaacctaac 720
agtttcctta agaaggctga tattttcatc aatgggtcgc agttaggaga agattttatt 780
caattgcata aacttctaag aaagtccacc ttcaaaaatg caaaactcta tggtcctgat 840
gttggtcagc ctcgaagaaa gacggctaag atgctgaaga gcttcctgaa ggctggtgga 900
gaagtgattg attcagttac atggcatcac tactatttga atggacggac tgctaccagg 960
gaagattttc taaaccctga tgtattggac atttttattt catctgtgca aaaagttttc 1020
caggtggttg agagcaccag gcctggcaag aaggtctggt taggagaaac aagctctgca 1080
tatggaggcg gagcgccctt gctatccgac acctttgcag ctggctttat gtggctggat 1140
aaattgggcc tgtcagcccg aatgggaata gaagtggtga tgaggcaagt attctttgga 1200
gcaggaaact accatttagt ggatgaaaac ttcgatcctt tacctgatta ttggctatct 1260
cttctgttca agaaattggt gggcaccaag gtgttaatgg caagcgtgca aggttcaaag 1320
agaaggaagc ttcgagtata ccttcattgc acaaacactg acaatccaag gtataaagaa 1380
ggagatttaa ctctgtatgc cataaacctc cataatgtca ccaagtactt gcggttaccc 1440
tatccttttt ctaacaagca agtggataaa taccttctaa gacctttggg acctcatgga 1500
ttactttcca aatctgtcca actcaatggt ctaactctaa agatggtgga tgatcaaacc 1560
ttgccacctt taatggaaaa acctctccgg ccaggaagtt cactgggctt gccagctttc 1620
tcatatagtt tttttgtgat aagaaatgcc aaagttgctg cttgcatc              1668
```

<210> 31
<211> 556
<212> PRT
<213> Artificial Sequence

<220>
<223> hep TEV110

<400> 31

```
Met Leu Leu Arg Ser Lys Pro Ala Leu Pro Pro Pro Leu Met Leu Leu
 1               5                  10                  15
Leu Leu Gly Pro Leu Gly Pro Leu Ser Pro Gly Ala Leu Pro Arg Pro
            20                  25                  30
Ala Gln Ala Gln Asp Val Val Asp Leu Asp Phe Phe Thr Gln Glu Pro
        35                  40                  45
Leu His Leu Val Ser Pro Ser Phe Leu Ser Val Thr Ile Asp Ala Asn
    50                  55                  60
Leu Ala Thr Asp Pro Arg Phe Leu Ile Leu Leu Gly Ser Pro Lys Leu
65                  70                  75                  80
Arg Thr Leu Ala Arg Gly Leu Ser Pro Ala Tyr Leu Arg Phe Gly Gly
                85                  90                  95
Thr Lys Thr Asp Phe Leu Ile Phe Asp Pro Lys Lys Glu Gly Ser Gly
            100                 105                 110
Ser Glu Asn Leu Tyr Phe Gln Gly Ser Gly Ser Thr Phe Glu Glu Arg
        115                 120                 125
Ser Tyr Trp Gln Ser Gln Val Asn Gln Asp Ile Cys Lys Tyr Gly Ser
    130                 135                 140
Ile Pro Pro Asp Val Glu Glu Lys Leu Arg Leu Glu Trp Pro Tyr Gln
145                 150                 155                 160
Glu Gln Leu Leu Leu Arg Glu His Tyr Gln Lys Lys Phe Lys Asn Ser
                165                 170                 175
Thr Tyr Ser Arg Ser Ser Val Asp Val Leu Tyr Thr Phe Ala Asn Cys
            180                 185                 190
Ser Gly Leu Asp Leu Ile Phe Gly Leu Asn Ala Leu Leu Arg Thr Ala
        195                 200                 205
Asp Leu Gln Trp Asn Ser Ser Asn Ala Gln Leu Leu Asp Tyr Cys
    210                 215                 220
Ser Ser Lys Gly Tyr Asn Ile Ser Trp Glu Leu Gly Asn Glu Pro Asn
225                 230                 235                 240
Ser Phe Leu Lys Lys Ala Asp Ile Phe Ile Asn Gly Ser Gln Leu Gly
                245                 250                 255
Glu Asp Phe Ile Gln Leu His Lys Leu Leu Arg Lys Ser Thr Phe Lys
            260                 265                 270
Asn Ala Lys Leu Tyr Gly Pro Asp Val Gly Gln Pro Arg Arg Lys Thr
        275                 280                 285
Ala Lys Met Leu Lys Ser Phe Leu Lys Ala Gly Gly Glu Val Ile Asp
    290                 295                 300
Ser Val Thr Trp His His Tyr Tyr Leu Asn Gly Arg Thr Ala Thr Arg
305                 310                 315                 320
Glu Asp Phe Leu Asn Pro Asp Val Leu Asp Ile Phe Ile Ser Ser Val
                325                 330                 335
Gln Lys Val Phe Gln Val Val Glu Ser Thr Arg Pro Gly Lys Lys Val
            340                 345                 350
Trp Leu Gly Glu Thr Ser Ser Ala Tyr Gly Gly Gly Ala Pro Leu Leu
        355                 360                 365
Ser Asp Thr Phe Ala Ala Gly Phe Met Trp Leu Asp Lys Leu Gly Leu
    370                 375                 380
Ser Ala Arg Met Gly Ile Glu Val Val Met Arg Gln Val Phe Phe Gly
385                 390                 395                 400
Ala Gly Asn Tyr His Leu Val Asp Glu Asn Phe Asp Pro Leu Pro Asp
                405                 410                 415
Tyr Trp Leu Ser Leu Leu Phe Lys Lys Leu Val Gly Thr Lys Val Leu
            420                 425                 430
Met Ala Ser Val Gln Gly Ser Lys Arg Arg Lys Leu Arg Val Tyr Leu
        435                 440                 445
His Cys Thr Asn Thr Asp Asn Pro Arg Tyr Lys Glu Gly Asp Leu Thr
    450                 455                 460
```

```
Leu Tyr Ala Ile Asn Leu His Asn Val Thr Lys Tyr Leu Arg Leu Pro
465                 470                 475                 480
Tyr Pro Phe Ser Asn Lys Gln Val Asp Lys Tyr Leu Leu Arg Pro Leu
                485                 490                 495
Gly Pro His Gly Leu Leu Ser Lys Ser Val Gln Leu Asn Gly Leu Thr
            500                 505                 510
Leu Lys Met Val Asp Asp Gln Thr Leu Pro Pro Leu Met Glu Lys Pro
        515                 520                 525
Leu Arg Pro Gly Ser Ser Leu Gly Leu Pro Ala Phe Ser Tyr Ser Phe
    530                 535                 540
Phe Val Ile Arg Asn Ala Lys Val Ala Ala Cys Ile
545                 550                 555
```

<210> 32
<211> 1710
<212> DNA
<213> Artificial Sequence

<220>
<223> hep TEV110/158

<400> 32

```
atgctgctgc gctcgaagcc tgcgctgccg ccgccgctga tgctgctgct cctggggccg 60
ctgggtcccc tctcccctgg cgccctgccc cgacctgcgc aagcacagga cgtcgtggac 120
ctggacttct tcacccagga gccgctgcac ctggtgagcc cctcgttcct gtccgtcacc 180
attgacgcca acctggccac ggacccgcgg ttcctcatcc tcctgggttc tccaaagctt 240
cgtaccttgg ccagaggctt gtctcctgcg tacctgaggt ttggtggcac caagacagac 300
ttcctaattt tcgatcccaa gaaggaaggc agcggatctg agaacctgta cttccagggt 360
tccggttcaa cctttgaaga gagaagttac tggcaatctc aagtcaacca ggatatttgc 420
aaatatggat ccatccctcc tgatgtggag gagaagttac ggttggaatg gccctaccag 480
gagcaattgc tactccgaga acactaccag tctggatccg gtgaaaatct ctattttcag 540
ggctcaggaa gtaaaaagtt caagaacagc acctactcaa gaagctctgt agatgtgcta 600
tacacttttg caaactgctc aggactggac ttgatctttg gcctaaatgc gttattaaga 660
acagcagatt tgcagtggaa cagttctaat gctcagttgc tcctggacta ctgctcttcc 720
aaggggtata acatttcttg ggaactaggc aatgaaccta acagtttcct taagaaggct 780
gatattttca tcaatgggtc gcagttagga gaagatttta ttcaattgca taaacttcta 840
agaaagtcca ccttcaaaaa tgcaaaactc tatggtcctg atgttggtca gcctcgaaga 900
aagacggcta agatgctgaa gagcttcctg aaggctggtg gagaagtgat tgattcagtt 960
acatggcatc actactattt gaatggacgg actgctacca gggaagattt tctaaaccct 1020
gatgtattgg acatttttat ttcatctgtg caaaaagttt tccaggtggt tgagagcacc 1080
aggcctggca agaaggtctg gttaggagaa acaagctctg catatggagg cggagcgccc 1140
ttgctatccg acacctttgc agctggcttt atgtggctgg ataaattggg cctgtcagcc 1200
cgaatgggaa tagaagtggt gatgaggcaa gtattctttg gagcaggaaa ctaccattta 1260
gtggatgaaa acttcgatcc tttacctgat tattggctat ctcttctgtt caagaaattg 1320
gtgggcacca aggtgttaat ggcaagcgtg caaggttcaa agagaaggaa gcttcgagta 1380
taccttcatt gcacaaacac tgacaatcca aggtataaag aaggagattt aactctgtat 1440
gccataaacc tccataatgt caccaagtac ttgcggttac cctatccttt ttctaacaag 1500
caagtggata aataccttct aagacctttg ggacctcatg gattactttc caaatctgtc 1560
caactcaatg gtctaactct aaagatggtg gatgatcaaa ccttgccacc tttaatggaa 1620
aaacctctcc ggccaggaag ttcactgggc ttgccagctt tctcatatag ttttttgtg 1680
ataagaaatg ccaaagttgc tgcttgcatc                                  1710
```

<210> 33
<211> 174
<212> PRT
<213> Homo sapiens

<400> 33

```
Gln Gln Asp Val Val Asp Leu Asp Phe Phe Thr Gln Glu Pro Leu His
 1               5                  10                  15
Leu Val Ser Pro Ser Phe Leu Ser Val Thr Ile Asp Ala Asn Leu Ala
```

```
            20                  25                  30
Thr Asp Pro Arg Phe Leu Ile Leu Leu Gly Ser Pro Lys Leu Arg Thr
        35                  40                  45
Leu Ala Arg Gly Leu Ser Pro Ala Tyr Leu Arg Phe Gly Gly Thr Lys
    50                  55                  60
Thr Asp Phe Leu Ile Phe Asp Pro Lys Lys Glu Ser Thr Phe Glu Glu
65                  70                  75                  80
Arg Ser Tyr Trp Gln Ser Gln Val Asn Gln Asp Ile Cys Lys Tyr Gly
                85                  90                  95
Ser Ile Pro Pro Asp Val Glu Glu Lys Leu Arg Leu Glu Trp Pro Tyr
            100                 105                 110
Gln Glu Gln Leu Leu Leu Arg Glu His Tyr Gln Lys Lys Phe Lys Asn
        115                 120                 125
Ser Thr Tyr Ser Arg Ser Ser Val Asp Val Leu Tyr Thr Phe Ala Asn
    130                 135                 140
Cys Ser Gly Leu Asp Leu Ile Phe Gly Leu Asn Ala Leu Leu Arg Thr
145                 150                 155                 160
Ala Asp Leu Gln Trp Asn Ser Ser Asn Ala Gln Leu Leu Leu
                165                 170
```

<210> 34
<211> 174
<212> PRT
<213> Mus musculus

<400> 34

```
Thr Asp Asp Val Val Asp Leu Glu Phe Tyr Thr Lys Arg Pro Leu Arg
 1               5                  10                  15
Ser Val Ser Pro Ser Phe Leu Ser Ile Thr Ile Asp Ala Ser Leu Ala
            20                  25                  30
Thr Asp Pro Arg Phe Leu Thr Phe Leu Gly Ser Pro Arg Leu Arg Ala
        35                  40                  45
Leu Ala Arg Gly Leu Ser Pro Ala Tyr Leu Arg Phe Gly Gly Thr Lys
    50                  55                  60
Thr Asp Phe Leu Ile Phe Asp Pro Asp Lys Glu Pro Thr Ser Glu Glu
65                  70                  75                  80
Arg Ser Tyr Trp Lys Ser Gln Val Asn His Asp Ile Cys Arg Ser Glu
                85                  90                  95
Pro Val Ser Ala Ala Val Leu Arg Lys Leu Gln Val Glu Trp Pro Phe
            100                 105                 110
Gln Glu Leu Leu Leu Leu Arg Glu Gln Tyr Gln Lys Glu Phe Lys Asn
        115                 120                 125
Ser Thr Tyr Ser Arg Ser Ser Val Asp Met Leu Tyr Ser Phe Ala Lys
    130                 135                 140
Cys Ser Gly Leu Asp Leu Ile Phe Gly Leu Asn Ala Leu Leu Arg Thr
145                 150                 155                 160
Pro Asp Leu Arg Trp Asn Ser Ser Asn Ala Gln Leu Leu Leu
                165                 170
```

<210> 35
<211> 174
<212> PRT
<213> Bos taurus

<400> 35

```
Ala Asp Asp Ala Ala Glu Leu Glu Phe Phe Thr Glu Arg Pro Leu His
 1               5                  10                  15
Leu Val Ser Pro Ala Phe Leu Ser Phe Thr Ile Asp Ala Asn Leu Ala
            20                  25                  30
Thr Asp Pro Arg Phe Phe Thr Phe Leu Gly Ser Ser Lys Leu Arg Thr
```

```
        35                  40                  45
Leu Ala Arg Gly Leu Ala Pro Ala Tyr Leu Arg Phe Gly Gly Asn Lys
    50                  55                  60
Gly Asp Phe Leu Ile Phe Asp Pro Lys Lys Glu Pro Ala Phe Glu Glu
65                  70                  75                  80
Arg Ser Tyr Trp Leu Ser Gln Ser Asn Gln Asp Ile Cys Lys Ser Gly
                85                  90                  95
Ser Ile Pro Ser Asp Val Glu Glu Lys Leu Arg Leu Glu Trp Pro Phe
            100                 105                 110
Gln Glu Gln Val Leu Leu Arg Glu Gln Tyr Gln Lys Lys Phe Thr Asn
        115                 120                 125
Ser Thr Tyr Ser Arg Ser Ser Val Asp Met Leu Tyr Thr Phe Ala Ser
    130                 135                 140
Cys Ser Gly Leu Asn Leu Ile Phe Gly Val Asn Ala Leu Leu Arg Thr
145                 150                 155                 160
Thr Asp Met His Trp Asp Ser Ser Asn Ala Gln Leu Leu Leu
                165                 170


<210> 36
<211> 173
<212> PRT
<213> Gallus gallus

<400> 36
Pro Arg Arg Thr Ala Glu Leu Gln Leu Gly Leu Arg Glu Pro Ile Gly
 1               5                  10                  15
Ala Val Ser Pro Ala Phe Leu Ser Leu Thr Leu Asp Ala Ser Leu Ala
            20                  25                  30
Arg Asp Pro Arg Phe Val Ala Leu Leu Arg His Pro Lys Leu His Thr
        35                  40                  45
Leu Ala Ser Gly Leu Ser Pro Gly Phe Leu Arg Phe Gly Gly Thr Ser
    50                  55                  60
Thr Asp Phe Leu Ile Phe Asn Pro Asn Lys Asp Ser Thr Trp Glu Glu
65                  70                  75                  80
Lys Val Leu Ser Glu Phe Gln Ala Lys Asp Val Cys Glu Ala Trp Pro
                85                  90                  95
Ser Phe Ala Val Val Pro Lys Leu Leu Leu Thr Gln Trp Pro Leu Gln
            100                 105                 110
Glu Lys Leu Leu Leu Ala Glu His Ser Trp Lys Lys His Lys Asn Thr
        115                 120                 125
Thr Ile Thr Arg Ser Thr Leu Asp Ile Leu His Thr Phe Ala Ser Ser
    130                 135                 140
Ser Gly Phe Arg Leu Val Phe Gly Leu Asn Ala Leu Leu Arg Arg Ala
145                 150                 155                 160
Gly Leu Gln Trp Asp Ser Ser Asn Ala Lys Gln Leu Leu
                165                 170


<210> 37
<211> 189
<212> PRT
<213> Homo sapiens

<400> 37
Glu Lys Thr Leu Ile Leu Leu Asp Val Ser Thr Lys Asn Pro Val Arg
 1               5                  10                  15
Thr Val Asn Glu Asn Phe Leu Ser Leu Gln Leu Asp Pro Ser Ile Ile
            20                  25                  30
His Asp Gly Trp Leu Asp Phe Leu Ser Ser Lys Arg Leu Val Thr Leu
        35                  40                  45
Ala Arg Gly Leu Ser Pro Ala Phe Leu Arg Phe Gly Gly Lys Arg Thr
```

```
    50                  55                  60
Asp Phe Leu Gln Phe Gln Asn Leu Arg Asn Pro Ala Lys Ser Arg Gly
65                  70                  75                  80
Gly Pro Gly Pro Asp Tyr Tyr Leu Lys Asn Tyr Glu Asp Asp Ile Val
                85                  90                  95
Arg Ser Asp Val Ala Leu Asp Lys Gln Lys Gly Cys Lys Ile Ala Gln
            100                 105                 110
His Pro Asp Val Met Leu Glu Leu Gln Arg Glu Lys Ala Ala Gln Met
        115                 120                 125
His Leu Val Leu Leu Lys Glu Gln Phe Ser Asn Thr Tyr Ser Asn Leu
    130                 135                 140
Ile Leu Thr Ala Arg Ser Leu Asp Lys Leu Tyr Asn Ser Ala Asp Cys
145                 150                 155                 160
Ser Gly Leu His Leu Ile Phe Ala Leu Asn Ala Leu Arg Arg Asn Pro
                165                 170                 175
Asn Asn Ser Trp Asn Ser Ser Ser Ala Leu Ser Leu Leu
            180                 185
```

<210> 38
<211> 151
<212> PRT
<213> Bombyx mori

<400> 38

```
Val Arg Tyr Phe Val Thr Ile Asn Glu Asn Gln Glu Asp Ile Lys Leu
 1               5                  10                  15
Ile Ser Glu Asp Phe Leu Ser Phe Gly Ile Asp Thr Ile Glu Ile Glu
            20                  25                  30
Asn Tyr Asn Arg Ile Asn Tyr Ser Asp Thr Arg Leu Arg Glu Leu Ala
        35                  40                  45
Ala Ala Leu Ser Pro Ala Arg Leu Arg Leu Gly Gly Thr Met Ser Glu
    50                  55                  60
Arg Leu Ile Phe Ser Lys Glu Asn Ile Pro Ile Ser Cys His Asn Cys
65                  70                  75                  80
Ser Tyr Lys Ser Tyr Pro Lys Ser Leu Cys Gln Leu Ile Glu Lys Pro
                85                  90                  95
Cys Lys His Lys His Lys Phe Leu Pro Phe Phe Ile Met Thr Gly Asn
            100                 105                 110
Glu Trp Asn Gln Ile Asn Asp Phe Cys Arg Lys Thr Asn Leu Lys Leu
        115                 120                 125
Leu Phe Ser Leu Asn Ala Met Leu Arg Asp Asn His Gly Trp Asn Glu
    130                 135                 140
Lys Asn Ala Arg Glu Leu Ile
145                 150
```

<210> 39
<211> 147
<212> PRT
<213> Hirudinaria manillensis

<400> 39

```
Lys Asn Val Ile Ala Ser Val Ser Glu Ser Phe His Gly Val Ala Phe
 1               5                  10                  15
Asp Ala Ser Leu Phe Ser Pro Lys Gly Pro Trp Ser Phe Val Asn Ile
            20                  25                  30
Thr Ser Pro Lys Leu Phe Lys Leu Leu Glu Gly Leu Ser Pro Gly Tyr
        35                  40                  45
Phe Arg Val Gly Gly Thr Phe Ala Asn Trp Leu Phe Phe Asp Leu Asp
    50                  55                  60
Glu Asn Asn Lys Trp Lys Asp Tyr Trp Ala Phe Lys Asp Lys Thr Pro
```

65 70 75 80
Glu Thr Ala Thr Ile Thr Arg Arg Trp Leu Phe Arg Lys Gln Asn Asn
85 90 95
Leu Lys Lys Glu Thr Phe Asp Asp Leu Val Lys Leu Thr Lys Gly Ser
100 105 110
Lys Met Arg Leu Leu Phe Asp Leu Asn Ala Glu Val Arg Thr Gly Tyr
115 120 125
Glu Ile Gly Lys Lys Thr Thr Ser Thr Trp Asp Ser Ser Glu Ala Glu
130 135 140
Lys Leu Phe
145

<210> 40
<211> 150
<212> PRT
<213> Scutellaria baicallensis

<400> 40
Asn Tyr Val Cys Ala Thr Leu Asp Leu Trp Pro Pro Thr Lys Cys Asn
1 5 10 15
Tyr Gly Asn Cys Pro Trp Gly Lys Ser Ser Phe Leu Asn Leu Asp Leu
20 25 30
Asn Asn Asn Ile Ile Arg Asn Ala Val Lys Glu Phe Ala Pro Leu Lys
35 40 45
Leu Arg Phe Gly Gly Thr Leu Gln Asp Arg Leu Val Tyr Gln Thr Ser
50 55 60
Arg Asp Glu Pro Cys Asp Ser Thr Phe Tyr Asn Asn Thr Asn Leu Ile
65 70 75 80
Leu Asp Phe Ser His Ala Cys Leu Ser Leu Asp Arg Trp Asp Glu Ile
85 90 95
Asn Gln Phe Ile Leu Glu Thr Gly Ser Glu Ala Val Phe Gly Leu Asn
100 105 110
Ala Leu Arg Gly Lys Thr Val Glu Ile Lys Gly Ile Ile Lys Asp Gly
115 120 125
Gln Tyr Leu Gly Glu Thr Thr Thr Ala Val Gly Glu Trp Asp Tyr Ser
130 135 140
Asn Ser Lys Phe Leu Ile
145 150

<210> 41
<211> 138
<212> PRT
<213> Arabidopsis thaliana

<400> 41
Asn Phe Val Cys Ala Thr Leu Asp Trp Trp Pro His Asp Lys Cys Asn
1 5 10 15
Tyr Asp Gln Cys Pro Trp Gly Tyr Ser Ser Val Ile Asn Met Asp Leu
20 25 30
Thr Arg Pro Leu Leu Thr Lys Ala Ile Lys Ala Phe Lys Pro Leu Arg
35 40 45
Ile Arg Ile Gly Gly Ser Leu Gln Asp Gln Val Ile Tyr Asp Val Gly
50 55 60
Asn Leu Lys Thr Pro Cys Arg Pro Phe Gln Lys Met Asn Ser Gly Leu
65 70 75 80
Phe Gly Phe Ser Lys Gly Cys Leu His Met Lys Arg Trp Asp Glu Leu
85 90 95
Asn Ser Phe Leu Thr Ala Thr Gly Ala Val Val Thr Phe Gly Leu Asn
100 105 110
Ala Leu Arg Gly Arg His Lys Leu Arg Gly Lys Ala Trp Gly Gly Ala

```
        115                 120                 125
Trp Asp His Ile Asn Thr Gln Asp Phe Leu
    130                 135
```

<210> 42
<211> 138
<212> PRT
<213> Arabidopsis thaliana

<400> 42

```
Asp Phe Ile Cys Ala Thr Leu Asp Trp Trp Pro Pro Glu Lys Cys Asp
 1               5                  10                  15
Tyr Gly Ser Cys Ser Trp Asp His Ala Ser Ile Leu Asn Leu Asp Leu
            20                  25                  30
Asn Asn Val Ile Leu Gln Asn Ala Ile Lys Ala Phe Ala Pro Leu Lys
        35                  40                  45
Ile Arg Ile Gly Gly Thr Leu Gln Asp Ile Val Ile Tyr Glu Thr Pro
    50                  55                  60
Asp Ser Lys Gln Pro Cys Leu Pro Phe Thr Lys Asn Ser Ser Ile Leu
65                  70                  75                  80
Phe Gly Tyr Thr Gln Gly Cys Leu Pro Met Arg Arg Trp Asp Glu Leu
                85                  90                  95
Asn Ala Phe Phe Arg Lys Thr Gly Thr Lys Val Ile Phe Gly Leu Asn
            100                 105                 110
Ala Leu Ser Gly Arg Ser Ile Lys Ser Asn Gly Glu Ala Ile Gly Ala
        115                 120                 125
Trp Asn Tyr Thr Asn Ala Glu Ser Phe Ile
    130                 135
```

<210> 43
<211> 138
<212> PRT
<213> Arabidopsis thaliana

<400> 43

```
Asn Phe Ile Cys Ala Thr Leu Asp Trp Trp Pro Pro Glu Lys Cys Asn
 1               5                  10                  15
Tyr Asp Gln Cys Pro Trp Gly Tyr Ala Ser Leu Ile Asn Leu Asn Leu
            20                  25                  30
Ala Ser Pro Leu Leu Ala Lys Ala Ile Gln Ala Phe Arg Thr Leu Arg
        35                  40                  45
Ile Arg Ile Gly Gly Ser Leu Gln Asp Gln Val Ile Tyr Asp Val Gly
    50                  55                  60
Asp Leu Lys Thr Pro Cys Thr Gln Phe Lys Lys Thr Asp Asp Gly Leu
65                  70                  75                  80
Phe Gly Phe Ser Glu Gly Cys Leu Tyr Met Lys Arg Trp Asp Glu Val
                85                  90                  95
Asn His Phe Phe Asn Ala Thr Gly Ala Ile Val Thr Phe Gly Leu Asn
            100                 105                 110
Ala Leu His Gly Arg Asn Lys Leu Asn Gly Thr Ala Trp Gly Gly Asp
        115                 120                 125
Trp Asp His Thr Asn Thr Gln Asp Phe Met
    130                 135
```